Physics Lib.

NEW SERVICE TO READERS -- SEE SECOND FOLLOWING BOOK REVIEWS

VOLUME 39 SECOND SERIES NUMBER 4

THE PHYSICAL REVIEW

A JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS

CONDUCTED BY

THE AMERICAN PHYSICAL SOCIETY

BOARD OF EDITORS

Managing Editor, JOHN T. TATE, University of Minnesota, Minneapolis
Assistant Editor, J. W. BUCHTA, University of Minnesota, Minneapolis
K. F. HERZFELD, A. W. HULL, F. W. LOOMIS, A. J. DEMPSTER, DAVID M. DENNISON, J. C. SLATER, R. C. GIBBS, E. O. LAWRENCE, J. H. VAN VLECK

FEBRUARY 15, 1932

Scattering of X-rays by Polyatomic Gases - - - - - - - - - - - Y. H. Woo
Note on Woo's Paper on the Scattering of X-rays by Polyatomic Gases - G. E. M. Jauncey
Measurements in the Long Wave-Length Infrared from 20μ to 135μ - R. Bowling Barnes
Direct Measurement of Mean Lives of Atomic States - - H. D. Kosins and A. Ellett
Inelastic and Elastic Electron Scattering in Argon - - A. L. Hughes and J. H. McMillen
New Determination of e/m From the Zeeman Effect - J. S. Campbell and W. V. Houston
Dirac's Equation and the Spin-Spin Interactions of Two Electrons - - - - G. Breit
Investigations in the Field of the Ultra-Short Electromagnetic Waves
I. The Generator For the Production of Ultra-Short Undamped Waves—G. Potapenko
Investigations in the Field of the Ultra-Short Electromagnetic Waves.
II. The Normal Waves and the Dwarf Waves - - - - - G. Potapenko
Determination of the Total Electric Polarization and the Electric Moments of Certain Organic Molecules - - - - - - - - - - - - - - - Louis M. Hunt
On the Stability of Certain Heteropolar Crystals - - - - - - - H. M. Evjen
Operating Characteristics of the Electro-Optical Shutter - - - Harold W. Wagner
New Kind of e.m.f. and Other Effects Thermodynamically Connected With the Four Transverse Effects - - - - - - - - - - - - - - - P. W. Bridgman
Electricity as a Natural Property of Riemannian Geometry - - - Cornelius Lanczos
Velocity of Sound in Tubes at Audible and Ultrasonic Frequencies - Charles B. Wales
Letters to the Editor - - - - - - - - - - - - - - - - - - -
Increasing the Charge Sensitivity of Western Tube Amplifiers - - - G. F. Mitchell
Effect of Heat on Mercury Resist - - - - - - - - - - J. Gordon Wilson
Resistant Electric Moment of Thorium Monoxide - - - - - - Henry Eyring
Another New X-ray Non-Diffracting Band - - - - - - - - - F. A. Kirkpatrick
Note on the New Effect Produced by Passing X-rays in Matter - - Heyrovsky and Grebenick
Book Reviews -
Current Literature of Physics

CALTECH LIBRARY

PUBLISHED BY THE AMERICAN PHYSICAL SOCIETY

THE AMERICAN PHYSICAL SOCIETY

OFFICERS OF THE SOCIETY

President:	W. F. G. Swann	Bartol Research Foundation, Philadelphia, Pa.
Vice-President:	Paul D. Foote	The Gulf Companies, Pittsburgh, Pa.
Secretary:	W. L. Severinghaus	Columbia University, New York City
Treasurer:	G. B. Pegram	Columbia University, New York City
Editor:	John T. Tate	University of Minnesota, Minneapolis, Minn.
Local Secretary for the Pacific Coast:	L. B. Loeb	University of California, Berkeley, Calif.

THE PHYSICAL REVIEW

Published by the American Physical Society. Executive office: University of Minnesota, Minneapolis, Minn. Two volumes annually, twelve issues per volume.

First series. From July 1, 1893, to January 1, 1913, the Review was conducted by Edward L. Nichols, Ernest Merritt and Frederick Bedell, with the financial support of Cornell University. The first four volumes were issued annually; subsequent volumes, semi-annually; thirty-five volumes in all.

Second series. On January 1, 1913, the Review was transferred to the American Physical Society as announced on the first page of the new series which was then begun. Since then it has been conducted by a Board of Editors elected by the Society. Semi-annual volumes begin in January and July. Since July 1, 1929, the Review has been issued twice a month.

Contributors are referred to page 2 of the advertisements for further information.

Annual subscription, throughout the Postal Union, $12.00 in advance.

General index, 1893-1920, $4.00 in advance.

Back Volumes, $9.00 per volume; Back Numbers, $1.50 per copy for issues prior to July 1, 1929 (Vols. 1-33), $.75 per copy for issues subsequent thereto. Postage is free if remittance accompanies order; otherwise postage is extra. No bound volumes can be supplied. The Review offers to purchase a limited number of First Series, Vol. 1, No. 1 (at $2.00 each); Second Series, Vol. 7, No. 1, January 1916; Vol. 7, No. 2, February 1916; Vol. 7, No. 4, April 1916; and Vol. 7, No. 5, May 1916 (at $2.00 each).

Advertisements. Rates will be supplied by the Editor on request. Copy for each number must be received at Minneapolis two weeks prior to the dates of issue; old copy is used until new copy is received. Send electrotypes direct to printers, George Banta Publishing Co., Menasha, Wis.

Discount of ten percent is allowed to members of the American Physical Society on purchase of back numbers.

Changes of Address and Complaints of failure to receive the Review, in the case of members of the Physical Society, should be addressed to the Treasurer; in the case of other subscribers, to the Editor. New copies can be sent free in response to complaints of non-delivery only if notice of non-delivery is received within three months of date of issue.

Editorial Correspondence should be addressed to the Physical Review, University of Minnesota, Minneapolis, Minn.

Entered at the post office at Menasha, Wis., as second class matter.
Acceptance for mailing at special rate of postage provided for in section 1103, Act of October 3, 1917, authorized September 6, 1926.

GALVANOMETER SHUNTS

(UNIVERSAL AYRTON-MATHER TYPE)

Flexibility in the use of direct-current galvanometers of high sensitivity demands the use of the well-known Ayrton-Mather "universal" type of shunt. In this shunt the multiplying power is independent of the galvanometer resistance.

General Radio makes its TYPE 229 Universal Galvanometer Shunts in two models: one with 1000 ohms and the other with 10,000 ohms resistance.

Write for Literature

GENERAL RADIO COMPANY

CAMBRIDGE A, ∴ MASSACHUSETTS

PHYSICAL REVIEW

PREPARATION OF PRELIMINARY ABSTRACTS

An article will not be accepted for publication in THE PHYSICAL REVIEW unless the manuscript is accompanied by an adequate abstract for publication at the beginning of the article. This abstract is intended to aid the reader by furnishing an index and a brief summary of the contents of the article. Besides serving these purposes it should also be suitable for reproduction in abstract journals so as to make it unnecessary for the editors of these journals to have another abstract prepared.

As *an index* it should be complete; all the subjects, major and minor, concerning which new information is presented, should each be given, with sufficient precision so that any reader can tell from the abstract whether the article contains anything of interest to him. The subject indexes of abstract journals are fundamental in reference work. These indexes are prepared exclusively from the abstracts and whatever is omitted from the abstracts cannot be included in the index and may thus be lost. The writer of an abstract should therefore feel himself under an important obligation to his scientific colleagues to make sure that the abstract is accurate and complete, at least as an index.

As *a summary* the abstract should give briefly the conclusions of the article, important advances in experimental technique and theory, and all numerical results of general interest that may be conveniently given including all that might belong in a handbook and table of contents. It should give all the information that readers who are not specialists in the particular field involved might desire to know about the article thus saving them the time and trouble in referring to the article itself. Experience has shown that in general the length of the abstract should be from four to eight percent of the length of the article.

THE PHYSICAL REVIEW, 1923-1925, contains many examples of adequate abstracts. Most of these contain paragraph titles and subtitles which indicate the subjects concerning which new information is given and it is required that authors include such subtitles when all the information contained in the article does not refer to the subject indicated by the title of the article. Such subtitles may be frequently avoided by rewording the title so as to make it more precise.

Second Series *February 15, 1932* *Vol. 39, No. 4*

THE
PHYSICAL REVIEW

THE SCATTERING OF X-RAYS BY POLYATOMIC GASES

By Y. H. Woo
Department of Physics
National Tsing Hua University, Peiping, China

(Received December 26, 1931)

Abstract

Recently Jauncey has shown that the factor F in Compton's formula for the scattering of x-rays by an atom should be the average atomic structure factor F' instead of the true atomic structure factor. Taking this into account, the general expression for the intensity of total scattering of x-rays by polyatomic gases previously deduced by the writer is modified replacing F by F'. The theory is applied to the scattering of x-rays by diatomic gases and the theoretical scattered intensity is actually compared with the absolute measurements made by Wollan on the scattering of Mo $K\alpha$ rays by H_2, N_2 and O_2. The values of F and F' are calculated from the Hartree field. On the whole Wollan's results support the new formula, though the old one represents in each case the general feature of the scattering curve. It is pointed out that, for the scattering by a diatomic gas of like atoms, Jauncey's theory of the scattering of x-rays by polyatomic molecules consisting of atoms of one kind gives results practically identical with those of the present theory, provided the change of wave-length due to the Compton effect is corrected for.

IN A recent paper[1] the writer has developed a general theory of the intensity of total scattering of x-rays by polyatomic gases on the assumption that, in considering the scattering of x-rays by a polyatomic molecule, only the coherent radiation from the different atoms in the molecule will interfere with each other according to the classical wave-principles, whereas the incoherent radiation will be simply added up. The mathematical formulation is based upon the theoretical investigation by A. H. Compton[2] and Raman[3] on the scattering of x-rays by a dynamic atom in which the electrons are regarded as a gas distributed in an enclosure surrounding the nucleus. Compton's formula may be written

$$I_\phi = I_e\left\{F^2 + \frac{Z - F^2/Z}{[1 + \gamma(1 - \cos\phi)]^3}\right\}, \tag{1}$$

where I_ϕ is the intensity scattered to a distance R in a direction ϕ with the incident radiation, I_e is the independent scattering from a single electron as

[1] Y. H. Woo, Proc. Nat. Acad. Sci. **17**, 467 (1931). This paper is referred to later as I.
[2] A. H. Compton, Phys. Rev. **35**, 925 (1930).
[3] C. V. Raman, Indian J. Phys. **3**, 357 (1928).

calculated by J. J. Thomson,[4] F is the atomic structure factor of the scattering atom, $\gamma = h/mc\lambda$, and Z, h, m, c and λ have their usual significance.

Recently Jauncey[5] has pointed out that the factor F in Eq. (1) should be a quantity F' (to be called later the average atomic structure factor) instead of the true atomic structure factor as defined in the theory of the reflection of x-rays by crystals. This comes out from the fact that Compton and Jauncey make different assumptions regarding the probability function that any one electron in the scattering atom shall lie between distances r and $r+dr$ from the nucleus. While Compton[6] assumes that this probability is the same for every electron in the atom, Jauncey[7] postulates that it is different for electrons in the different electron groups. According to Jauncey,[7] the relation between the average atomic structure factor F' and the true atomic structure factor F is expressed by

$$F'^2 = F^2 - \left(Z \sum_{r=1}^{Z} E_r^2 - F^2 \right)/(Z - 1), \tag{2}$$

where E_r is the average amplitude of the waves scattered by the r-th electron in the scattering atom and where the true atomic structure factor F is given by

$$F = \sum_{1}^{Z} E_r. \tag{3}$$

Introducing the modification discussed by Jauncey, Compton's formula (I) becomes

$$I_\phi = I_e \left\{ F'^2 + \frac{Z - F'^2/Z}{[1 + \gamma(1 - \cos\phi)]^3} \right\}. \tag{4}$$

Thus, according to our fundamental assumption stated above, the intensity scattered at an angle ϕ with the primary beam by a molecule containing atoms $1, \ldots i, \ldots n$ arranged at fixed distances from each other is given by

$$I_{\phi n} = I_e \sum_{1}^{n} \sum_{1}^{n} F_i' F_j' \frac{\sin k s_{ij}}{k s_{ij}} + I_e \sum_{1}^{n} \frac{Z_i - F_i'^2/Z_i}{[1 + \gamma(1 - \cos\phi)]^3}, \tag{5}$$

where F_i' is the average atomic structure factor for the i^{th} atom, $k = (4\pi/\lambda) \sin\frac{1}{2}\phi$, s_{ij} is the distance of any atom i from any other atom j, and the other notations have the same meaning as those employed in Eq. (1). The expression (5) differs from the formula (2) of I in replacing F by F' as suggested by Eq. (4).

Now let us apply the above theory to the scattering of x-rays by a diatomic gas consisting of two like atoms. Since there is only one value of F' or F, we have for the scattering per electron

[4] J. J. Thomson, "Conduction of Electricity through Gases," 2nd Edition, p. 325.

[5] G. E. M. Jauncey, Phys. Rev. **37**, 1193 (1931). Cf. also G. Herzog, Zeits. f. Physik **69**, 207 (1931).

[6] A. H. Compton, reference 2, p. 930.

[7] G. E. M. Jauncey, reference 5, p. 1201.

$$S = \frac{I_{\phi 2}}{2ZI_e} = \left(1 + \frac{\sin ks}{ks}\right)\frac{A^2}{Z} + \frac{1 - A^2/Z^2}{[1 + \gamma(1 - \cos\phi)]^3}, \qquad (6)$$

where s is the distance between the two atoms in the scattering molecule, A is equal to F' according to Eq. (5) and is equal to F if Eq. (2) of I is followed.

Wollan[8] has recently reported experiments on the absolute measurements of the intensity of the scattering of Mo $K\alpha$ rays by H_2, N_2 and O_2. In order to compare Eq. (6) with Wollan's results, it is necessary to calculate the values of F and F' for the atoms in the scattering molecule. From the tables given by James and Brindley[9] or by the interpolation method designed by these authors,[9] we can easily find the contribution of a single electron of each of the electron groups in any of the three atoms H, N and O to its F. That is, we can find the values of the E_r's calculated according to the Hartree method. Thus the average atomic structure factor F' and the true atomic structure factor F can be evaluated according to Eqs. (2) and (3) without difficulty. For the case of hydrogen, F' is equal to F. For oxygen and nitrogen, we group the electrons into groups as shown in Table I and denote each group by the

TABLE I.

	Z	K (1 0)	L (2 0)	(2 1)
O	8	2	2	6
N	7	2	2	5

TABLE II. *Values of F and F'.*

ϕ	$\sin\frac{1}{2}\phi/\lambda$	H	N		O	
		$F=F'$	F	F'	F	F'
10°	0.123	0.72	5.53	5.52	6.75	6.74
15°	0.184	0.53	4.35	4.30	5.57	5.55
20°	0.245	0.36	3.53	3.44	4.74	4.69
25°	0.306	0.25	2.88	2.71	3.88	3.80
30°	0.365	0.16	2.43	2.18	3.22	3.07
40°	0.482	0.08	1.88	1.51	2.35	2.09
50°	0.596	0.04	1.67	1.27	1.88	1.50
60°	0.704	0.03	1.56	1.19	1.65	1.25
70°	0.808	0.02	1.46	1.11	1.56	1.18
80°	0.905	0.01	1.35	1.03	1.47	1.11
90°	0.996	0.00	1.26	0.96	1.39	1.05

quantum numbers (nl) in the usual way. The values of F' and F for H, N and O so obtained are tabulated in Table II. It will be noticed that for comparatively large values of $\sin\phi/2/\lambda$ the difference F—F' becomes quite appreciable. This is in agreement with the conclusion recently drawn by Jauncey.[10] In view of the fact, however, that the values of E_r employed in the present calculation are estimated from the Hartree field,[11] the values of F and F'

[8] E. O. Wollan, Phys. Rev. **37**, 862 (1931); Proc. Nat. Acad. Sci. **17**, 475 (1931).
[9] James and Brindley, Phil. Mag. **12**, 81 (1931).
[10] G. E. M. Jauncey, Phys. Rev. **38**, 1 (1931).
[11] Cf. James and Brindley, reference 9.

given here are perhaps more accurate than those obtained by the simple method devised by Jauncey.[10]

Using the values of F and F' tabulated in Table II, a calculation is made of the scattering of Mo $K\alpha$ rays by H_2, N_2 and O_2 according to Eq. (6). The values of s are taken to be equal to 1.10A, 1.21A and 1.10A for H_2, O_2 and N_2 respectively.[12] The theoretical values of the scattering per electron S together with the experimental results observed by Wollan[13] are given in Table III. In order to see the variation of S with the scattering angle ϕ, these results are

TABLE III. *Values of the scattering per electron.*

ϕ	H_2		N_2			O_2		
	S $(F=F')$	S obs.	S $(A=F)$	S $(A=F')$	S obs.	S $(A=F)$	S $(A=F')$	S obs.
10°	1.31	—	7.32	7.29	6.90	8.93	8.92	8.05
15°	1.11	1.27	3.90	3.84	—	4.89	4.86	4.90
20°	0.99	1.10	2.40	2.34	2.45	3.05	3.00	2.96
25°	0.98	1.03	1.76	1.67	—	2.24	2.19	2.00
30°	0.98	1.00	1.55	1.45	1.35	1.96	1.88	1.72
40°	0.98	0.98	1.44	1.27	1.25	1.67	1.52	1.43
50°	0.97	0.97	1.35	1.18	1.15	1.37	1.23	1.15
60°	0.95	0.95	1.24	1.12	1.05	1.22	1.10	1.04
70°	0.94	0.93	1.18	1.07	1.03	1.20	1.09	1.02
80°	0.92	0.92	1.15	1.05	1.01	1.17	1.07	1.01
90°	0.91	0.91	1.12	1.03	0.96	1.12	1.03	0.96

plotted in Figs. 1 and 2. The curve I assumes $A=F$ and the curve I' $A=F'$. The circles show the observed values. It is seen that on the whole Wollan's experiments support Eq. (6) for the case $A=F'$, though Eq. (6) with $A=F$ represents in each case the general feature of the scattering curve.

In this connection it may be mentioned that Wollan[14] has previously made a comparison of his own data with Eq. (6) for the case $A=F$. He estimated the F values from the atomic field of Thomas[15] and Fermi.[16] The agreement between theory and experiment is in general not quite satisfactory. This is perhaps well explained by the consideration just given above. Assuming $A=F$, the writer[17] has also compared Eq. (6) with the experiments reported by Barrett[18] on the scattering of x-rays of various wave-lengths by diatomic gases. In some cases the agreement seems to be all right. However, since Barrett's measurements give relative values of the scattering per electron, but not absolute values, there may remain some arbitrariness in fitting the observed data with the calculated results. So far as the writer is aware, the experiments of Wollan seem to be the only ones which give absolute val-

[12] These values of s are calulated according to Rasetti's measurements of the moment of inertia, cf. Phys. Rev. **34,** 367 (1930)

[13] E. O. Wollan, reference 8.

[14] E. O. Wollan, Proc. Nat. Acad. Sci. **17,** 475 (1931).

[15] L. H. Thomas, Proc. Camb. Phil. Soc. **23,** 542 (1927).

[16] E. Fermi, Zeits. f. Physik **48,** 73 (1928).

[17] Y. H. Woo, Proc. Nat. Acad. Sci. **17,** 467 and 470 (1931).

[18] C. S. Barrett, Proc. Nat. Acad. Sci. **14,** 20 (1928); Phys. Rev. **34,** 22 (1928).

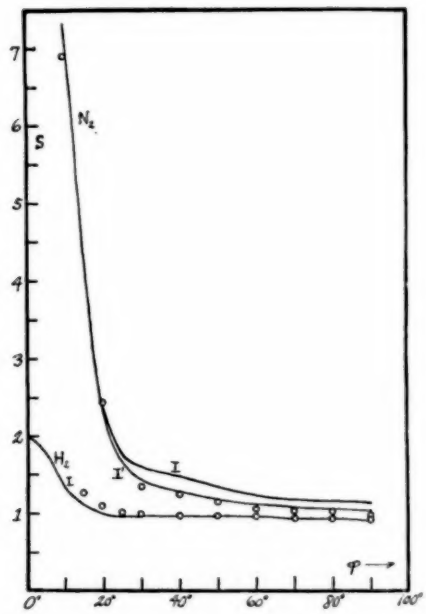


Fig. 1. Curve *I*, $A = F$; Curve I', $A = F'$; circles, experimental points.

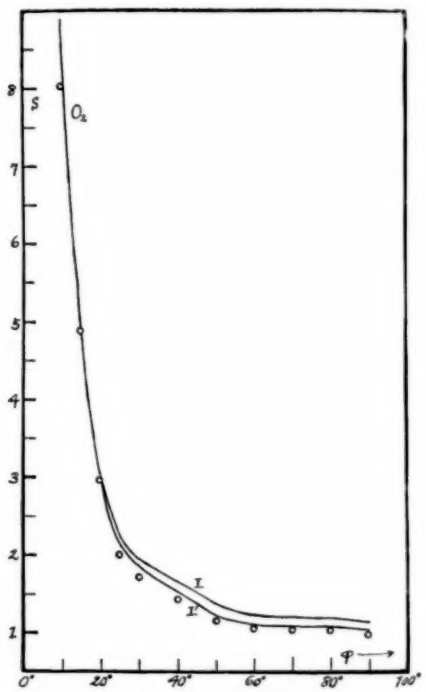


Fig. 2. Curve I, $A = F$; Curve I', $A = F'$; circles, experimental points.

ues. This shows that absolute measurements on the intensity of total scattering of x-rays by polyatomic gases and vapors for different wave-lengths are much desired.

As is well known, the factor $[1+\gamma(1-\cos\phi)]^{-3}$ is introduced to correct for the change of wave-length in the Compton effect for the incoherent part of the intensity of the total scattering. For the case of the scattering of x-rays by monatomic gases, its importance has been emphasized by the writer.[19] Though the maximum scattering angle in Wollan's experiments only amounts to 90°, the significance of this factor is well illustrated by the following example. In Table IV are given the values of the scattering per electron for the

TABLE IV. *Uncorrected values of scattering per electron for* O_2.

	10°	15°	20°	25°	30°	40°	50°	60°	70°	80°	90°
S (uncorrected)	8.92	4.87	3.01	2.20	1.89	1.54	1.26	1.15	1.15	1.14	1.13

scattering of Mo $K\alpha$ by O_2 calculated according to Eq. (6) for $A = F'$ without the correction factor. In comparison with the corresponding corrected and observed values of S listed in Table III, it is evident that the corrected values are in better accord with the experiment. Similar results are obtained for the scattering of Mo $K\alpha$ rays by H_2 and N_2.

Recently Jauncey[20] has also proposed a theory of the scattering of x-rays by a polyatomic molecule consisting of atoms of one kind. When applied to the case of the scattering of x-rays by a diatomic gas of two like atoms, Jauncey's formula may be written

$$S = \frac{F'^2}{Z} + \frac{\sin ks}{ks}\frac{F^2}{Z} + 1 - \frac{F'^2}{Z^2}, \tag{7}$$

where the effect of the change of wave-length is neglected. When this effect is corrected for, Eq. (7) gives results practically identical with those calculated according so Eq. (6) for the case $A = F'$.

[19] Y. H. Woo, Proc. Nat. Acad. Sci. **16,** 814 (1930) and **17,** 470 (1931).
[20] G. E. M. Jauncey, Phys. Rev. **38,** 194 (1931).

NOTE ON WOO'S PAPER ON THE SCATTERING OF X-RAYS BY POLYATOMIC GASES

By G. E. M. Jauncey
Washington University, St. Louis, Missouri

(Received January 2, 1932)

WOO'S formulas (5) and (6) do not seem to me to be quite correct. When account is taken of the distinction between F and F' (that is, between f and f' in the notation used in the papers by Harvey and myself), formula (5) should read

$$I_{\varphi n} = I_e \sum_1^n F_i'^2 + I_e \sum_1^n{}' \sum_1^n{}' F_i F_j \frac{\sin ks_{ij}}{ks_{ij}} + I_e \sum_1^n \frac{Z_i - F_i'^2/Z_i}{[1+\gamma(1-\cos\phi)]^3} \quad (5a)$$

where the symbol $\Sigma'\Sigma'$ indicates that in the summation $i \neq j$. As a result of this change in Eq. (5), Woo's formula (6) for diatomic gas molecules consisting of two like atoms becomes

$$S = \frac{I_{\varphi 2}}{2ZI_e} = \frac{A^2}{Z} + \frac{\sin ks}{ks}\cdot\frac{F^2}{Z} + \frac{1 - A^2/Z^2}{[1+\gamma(1-\cos\phi)]^3}. \quad (6a)$$

The difference between Woo's formula and my formula (7) in his paper is that his formula takes account of the Compton change of wave-length while mine does not. Hence Woo's formula should reduce to mine when $\gamma = 0$. It is seen that formula (6a) does reduce to my formula (7) in Woo's paper when $A = F'$ and $\gamma = 0$. As a result of using Eq. (6a) instead of Woo's Eq. (6), some of the values in the columns headed $A = F'$ for N_2 and O_2 in Woo's Table III are slightly changed. The greatest change, however, is less than one third of the difference between the values given in the columns headed $A = F$ and $A = F'$, so that Woo's conclusions are not affected by this slight change.

MEASUREMENTS IN THE LONG WAVE-LENGTH INFRARED FROM 20μ TO 135μ

By R. Bowling Barnes*
Physikalisches Institut der Universität Berlin

(Received December 21, 1931)

Abstract

The technique employed in this spectral region is discussed, special emphasis being placed upon the problem of the elimination of shorter wave-lengths. New measurements from 20μ to 135μ, which show the transmission of 11 of the substances most commonly used in the infrared, are presented and discussed. In the light of new absorptions found in crystalline quartz, the reststrahlen measurements from Rubens are discussed. The transmission of sulphur was measured, and between 20μ and 135μ, 8 absorption bands were located. These are compared with the Raman effect measurements of Krishnamurti. The question of an ideal substance for coating receiving elements in the far infrared is also treated.

THE infrared spectrum may now be examined by means of prisms from the visible region out to about 28μ.[1,2] Here, at a rather sharp boundary, prism spectroscopy comes to an end, and in order to investigate regions of greater wave-length an entirely different procedure must be followed. The two regions created by this boundary line are known as the "near" infrared and the "far" infrared. It is the purpose of this paper to present some new measurments made in the far infrared, and to discuss the technique and methods employed in investigating this part of the spectrum. The experimental difficulties increase rather rapidly as one proceeds towards longer wave-lengths. Probably for this reason, work in the far infrared has been limited. The problems which have been successfully studied are so widely scattered and so varied in subject matter, that it is rather difficult for one to obtain a clear view of the spectroscopy of this region. Accordingly, it is thought that a few remarks in this regard may not be out of place.

Our first knowledge of this part of the spectrum came from the classical experiments of Rubens and his co-workers, who, using radiations monochromized by means of the reststrahlen method,[3] were able to study the optical behavior of various materials with respect to the far infrared wave-lengths. The absorption and reflection curves which were obtained by them have furnished preliminary information of untold value to all subsequent investigators of this region.

* National Research Fellow.

[1] Pohl in Göttingen has grown crystals of KBr large enough to yield prisms of 4 to 6 cm tall. The dispersion of KBr is known now to 28.5μ. See Gundelach, Zeits. f. Physik **66**, 775 (1930).

[2] J. Strong, Phys. Rev. **36**, 1663 (1930). Strong has grown larger single crystals of KBr, which he says are transparent to 30μ. Crystals of KI are transparent to 33μ.

[3] For details of this method and these measurements see Schaefer and Matossi, "Das ultrarote Spektrum".

It was soon realized however, that, for studying the detailed absorption or reflection of various substances, the dispersion obtained using the reststrahlen method was not great enough. The radiations so obtained were too impure, and their wave-lengths, which were really only average wave-lengths, depended too much upon such outside factors as atmospheric absorption and the condition of polarization of the original radiation. Accordingly, a new method was brought into use.[4] In this a wire grating was employed as the dispersing medium, the formerly used effect of selective reflection playing only a secondary role. This method, with a few minor changes, is today the accepted means of investigating the far infrared, and accordingly will be discussed briefly below.

Apparatus and Methods

The apparatus used in this present work is essentially that used by Czerny in most of his investigations and is shown schematically in Fig. 1. The radiation from the Welsbach mantle, W, passed through a transparent shutter, H, which consisted of a 3.0 mm fluorspar plate, and through an opaque layer of

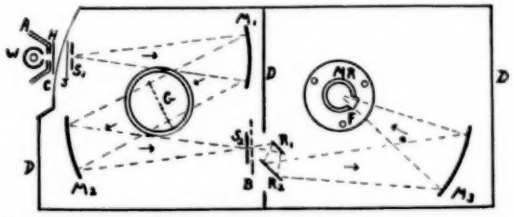


Fig. 1. Sketch of apparatus. W = Welsbach mantle, A = Aluminum shields, H = Fluorspar shutter, C = Celluloid window, D = Galvanized iron case, J = Opaque soot film, S_1, S_3 = Spectrometer slits, M_1, M_2, M_3 = Concave mirrors, G = 0.2 mm wire grating, B = Samples to be measured, R_1, R_2 = Reststrahlen plates, F = Window (quartz or paraffin), MR = Microradiometer.

camphor soot, J, before entering the first slit of the spectrometer at S_1. By the aid of the soot filter and the fluorspar shutter the disturbing effect of the short wave-lengths was greatly weakened. The mirrors M_1 and M_2 were each of 44 cm focal length and 10 cm diameter. By means of them the radiation emerging from S_1 was collimated, passed through the grating, G, and focussed again upon S_2. After leaving the spectrometer the radiation then passed through one of four openings in the brass frame B. Three of these openings were covered with samples to be studied, and the fourth left clear. At R_1 and R_2 were either two reststrahlen plates, or one plate and one silver mirror as the case necessitated. By means of M_3 a final image of S_1 was projected upon the thermoelement of a microradiometer, MR. For wave-lengths under 40μ the window of the latter was a 2.0 mm sheet of paraffin, and for $\lambda > 40\mu$, a plate of quartz 0,4 mm thick cut perpendicular to the axis. The entire apparatus, excepting the source and shutter, was enclosed in an airtight galvanized iron case, and the air dried as thoroughly as possible with P_2O_5. With the microradiometer mounted upon a Julius suspension, which made use of an

[4] H. Rubens, Berl. Berichte 1921, p. 8; H. Witt, Zeits. f. Physik **28,** 236, (1924); M. Czerny, Zeits. f. Physik **34,** 227 (1925); **44,** 235 (1927).

oil tank damping scheme recently described by Müller,[5] the zero, even with the scale at 5 m distance, was so steady that measurements could easily be made in the middle of the afternoon. The grating remained fixed, and the spectrometer arm carrying the source, shutter, first slit and M_1 was rotated about an axis in G in order to pass the various wave-lengths successively through S_2.

Since from a radiating body having the temperature of a Welsbach mantle, approximately 1800°K, the ratio of the total energy falling in the near infrared to that in the far infrared is so extremely large, one must exercise a great deal of care in eliminating all stray or false radiation. A very small percent of the 20μ energy, for example, occurring as an impurity could easily produce a deflection of the same size as that produced by 100 percent of the

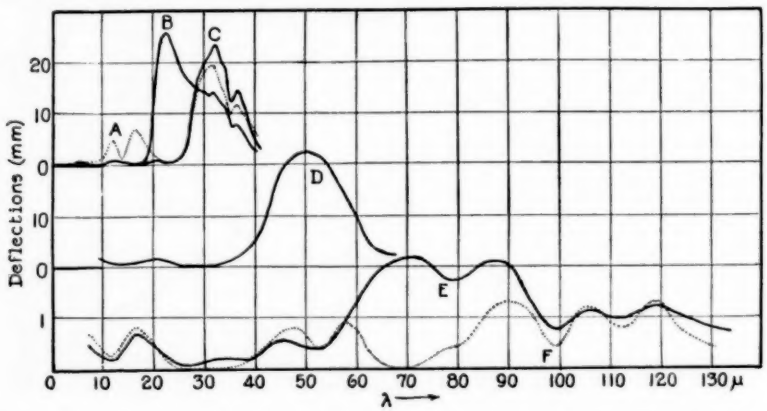


Fig. 2. Energy curves showing distribution of energy used in the various parts of the spectrum. A=0.2 grating, 2.5 mm slits, 2 mm paraffin, 1×Aragonite reflection. B=0.2 gratings, 2.5 mm slits, 2 mm paraffin, 2×CaF_2 reflections. C=0.2 gratings, 2.5 mm slits, 2 mm paraffin, 1×Aragonite×CaF_2 reflection. D=0.8 grating, 3.4 mm slits, 0.4 mm quartz (⊥) 1×NaCl reflection. E=0.8 grating, 3.4 mm slits, 0.4 mm quartz (⊥) 1 mm (CaF_2−paraffin), 1×TlCl. F=0.8 grating, 3.4 mm slits, 0.4 mm quartz (⊥) 1×TlBr. In all cases, Welsbach mantle, microradiometer, scale distance of 5 m, spectrometer f=4.4.

energy at 100μ. This danger is greatly increased because of the fact that gratings superpose orders. While, on account of the spacing chosen, the wire gratings in most common use automatically eliminate the even orders, for the measurements at 100μ the third order of 33μ, the fifth order of 20μ, etc., are dangerous. The spacings of such a grating are always somewhat nonuniform, and therefore, in actual practice, traces of the even orders may be present along with the odd orders. Then too, Rubens showed that for $<5\mu$ the wires themselves may act as cylindrical mirrors and reflect these short wave-lengths in the same direction as the diffracted 100μ radiation. It follows therefore, that through some method of filtering, either by selective absorption or selective reflection, all wave-lengths shorter than half those to be used should be entirely eliminated. As yet, no ideal filter substance has been found which will meet the requirements stated above. Fig. 2 shows the extent to which

[5] Müller, Ann. d. Physik **1**, 613 (1929).

these conditions have been fulfilled through the use of a combination of filters. The actual energies present, as are shown by the microradiometer deflections, are plotted as ordinates against the wave-lengths as abscissae. Attention is called to the increase of energy as one approaches the near infrared. This increase necessitates a change in the scale of the ordinates to 10 times that used at the longer wave-lengths. Curves *A*, *B* and *C* were obtained under the following conditions; Welsbach mantle, opaque camphor soot film deposited upon a 1μ celluloid film, 2.0 mm paraffin window, 0.2 mm grating, 2.5 mm slits, and reflections from 1×Aragonite, 2×CaF_2, and 1×Aragonite×CaF_2 respectively. Due to the presence of the two short wave-length maxima in the reflection power of Aragonite, curve *A* could not be used. These disturbing wave-lengths were easily cut out by introducing the reflection on CaF_2, giving curve *C*. With the 0.2 mm grating the slits were used 2.5 mm wide. This was equivalent to 20′ or 1.16μ. Curves *D*, *E* and *F* were made with Welsbach mantle, soot film as above, 0.4 mm quartz window, 0.8 mm grating, 3.4 mm slits and reflections from 1×NaCl, 1×TlCl, and 1×TlBr. In *E* and *F* it was necessary, because of the presence of too much energy between 30μ and 40μ, to introduce a 1 mm filter of paraffin containing 1 part in 10 of powdered CaF_2.[6] With this grating, 0.8 mm, the slit width used was equivalent to 27′ or 6.2μ. Realizing that the scale was 5 m distant from the microradiometer, the size of these deflections shows at a glance the difficulty of measuring percentage transmissions or reflections in this part of the spectrum.

Only the most sensitive receiving instruments may be used, and all outside effects such as those arising from building vibrations, sudden changes of the room temperature, etc., must be carefully controlled.[7] Because the deflections were so small, their exact endpoints were hard to determine. The following method was therefore adopted. With larger deflections, it was determined that the time necessary for the deflection to reach its maximum value was 16 seconds. A metronome was used, and regardless of the size of the deflection, the shutter was closed at the end of the first 16 seconds and the deflection noted. At the end of the second 16 seconds, the zero point was again read and due allowance made for any drift that had taken place. With the frame *B* set so as to present the clear opening to S_2, 2 or 3 deflections were measured. Then, with one of the samples before S_2, 5 or 6 deflections were read, after which, 2 or 3 more without the sample were measured. From the average values with and without the sample in the path of the radiation, the percentage transmission was obtained. At some wave-lengths where the deflections were exceptionally small, 8 or 10 values were used in making up the averages. Even then, in some cases the results were more of a qualitative than of a quantitative nature.

[6] M. Czerny, Zeits. f. Physik **53,** 317 (1929). Since the powdered CaF_2 settles to the bottom of the molten paraffin, the exact constituency of this filter is uncertain.

[7] In order to obtain constant temperature conditions, the measuring room was closed and the Welsbach mantle lighted at least three hours before beginning each series of measurements. Then, upon entering the room the observer took his place and waited still another half hour before beginning. On days when the weather changed rapidly, measurements could not be made, because the zero of the microradiometer changed rapidly also.

It should be remarked here, that since no ideal distribution of energy can be obtained, and since the slits must be rather wide, that, in regions where the absorption of the sample which is being measured changes rapidly, the slit-width effects are very noticeable. In these cases the Paschen-Runge slit-width correction[8] should be made in order to obtain the true form of the measured curve.

Results

Most of the substances discussed below have from time to time been used in some manner in the infrared. Up to the present, however, no series of connected measurements upon them have been reported, the existing values consisting of a few measured points scattered here and there throughout the spectrum. Since from these existing measurements no complete view of the transmission of these substances could be obtained, the following measurements were undertaken, and the results there from are given in the curves which are discussed below. The percentage transmissions, irrespective of whether the loss is due to true absorption, reflection or scattering, is in every case plotted as a function of the wave-length. These measurements extend from 135μ to 20μ, but wherever possible, the curves have been extended below 20μ. In these cases measurements are taken from other authors, and are given here merely for the sake of better orientation and in order to present a more complete picture of the behavior of the substances in question. Since the main theme of this paper is the far infrared, the abscissae have been plotted to a scale in μ itself instead of log μ or ν. This is, of course, not fair to the near infrared, and therefore I wish to express here my apologies.

Fused quartz, Fig. 3

In the near infrared this substance absorbs very strongly. In fact one reads in the literature, that, "beyond 11.6μ fused quartz is non-transparent."

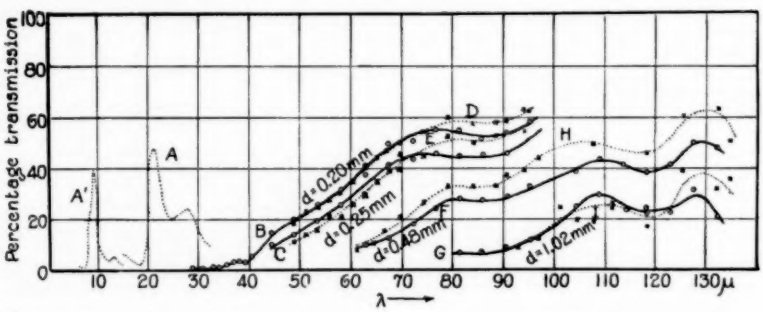


Fig. 3. Transmission of fused quartz. A' and A are reflection power curves. The others, percentage transmission as function of λ.

Because of this exceptionally strong absorption, the near infrared is represented in this figure by the reflection curves A' and A from Schaefer and Rubens respectively. Curves B, C, F and G show the transmission of samples having thicknesses equal to 0.20 mm, 0.25 mm, 0.48 mm and 1.02 mm. Be-

[8] "Das ultrarote Spektrum", p. 80.

sides the general absorption for wave-lengths under 40μ, absorption bands exist in neighborhoods of 86μ, 118μ and $\lambda > 132\mu$. Though these bands are broad and rather weak, they indicate in all probability the presence of small groups of recrystallized quartz molecules.[9] The fact that the four samples which were measured were obtained from three separate sources where they undoubtedly underwent different annealing treatments, seems to show that the presence of these recrystallized regions is a property of fused quartz in general.[10]

Since throughout most of the above measurements the deflections were very small, of the order of 1 to 2 mm, it was thought best to remeasure the absorption of fused quartz with entirely different conditions. This was, of course, to be a test for the presence in the curves of systemmatic errors due to some fault of the apparatus, or to false radiation, etc. With entirely different filter combinations, the dotted curves *D*, *E*, *H* and *I* were measured. The agreement between these curves and the ones described above shows that the absorption bands are real, and at the same time presents an idea of the accuracy, which was obtained in these new measurements in spite of the small deflections which had to be used. In filtering out the short wave-lengths the following selective reflections were substituted for those used in obtaining the respective energy curves of Fig. 2: from 49μ to 55μ 1×NaCl, from 55μ to 70μ 1×KCl, from 70μ to 94μ 1×KBr and from 94μ to 134μ 2×TlCl.

Crystalline quartz, Fig. 4

Again, for similar reasons as above, a reflection power curve from Rubens is given in the near infrared. From it, curve *A*, we can read off indicated posi-

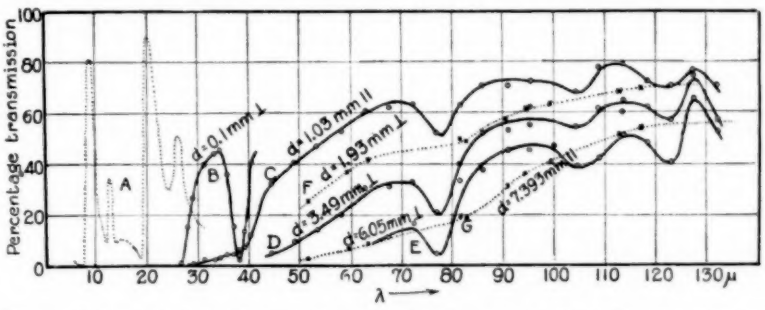


Fig. 4. Transmission of crystalline quartz. *A*, reflection curve. *B*, *C*, *D*, and *E*, transmission curves. *F* and *G*, transmission curves measured with the reststrahlen method.

tions of rather intense absorption at 8.5μ, 12.5μ, 20.5μ and 26μ. Curve *B* was made with a piece of quartz 0.10 mm cut perpendicular to the optic axis. After increasing in transparency, following the 26μ absorption, it shows clearly the previously known 38μ[11] band. Curves *C* (1.03 mm ||), *D*(3.49 mm ⊥), and

[9] The very fact that these bands lie at wave-lengths greater than say 30μ, shows, according to the accepted theory of absorption in solids, that they represent frequencies characteristic of a crystal lattice.

[10] O. Reinkober, Phys. Zeits. **32,** 243 (1931). In examining the tensil strength of quartz fibers Reinkober always found centers of recrystallized quartz present.

[11] See reference 6.

E(6.05 mm $\perp$) show definitely bands at 77μ,[11] 105μ and 122. The last point at 132μ indicates the presence of still another band at longer wave-lengths. Towards longer wave-lengths the reflection power of quartz increases. Therefore, because of the increased losses of energy due to the reflections at the two surfaces of the samples, the transmission curves, especially C, after about 90μ begin to run almost horizontally. No effect of interference due to these reflections however is to be seen. Attention is called to the fact that curve C (1.0 mm 11) shows definitely at 38μ a transparency several times higher than that of curve B (0.10 mm $\perp$). This is probably an effect of dichroismus.[12]

In the case of crystalline quartz a test was also made to discover the effects, if any, arising from false radiation. With the energy curve F in Fig.2 (1$\times$TlBr) the absorption of curve E from 95μ to 118μ was remeasured. The values agreed to within 2 percent with those of curve E, giving the location of the band again at 105μ.

The two dotted curves, F(1.93 mm $\perp$) and G(7.393 mm 11) are from Rubens[13] and show no definite evidence of absorption bands. This failure of absorption bands to show up in these curves throws a very interesting sidelight upon the method previously used in this region, namely the reststrahlen method. It seems more than a mere accident that the measured points always fell a little to one side or to the other of the quartz bands themselves. It will be noticed, that no one of the 9 measured points occurred at a wave-length which the quartz absorbed strongly. The following explanation of this fact is offered as a possibility.

Since the wave-lengths of the various reststrahlen used in the above curves F and G were determined by means of a quartz interferometer, they of course depended to a great extent upon the assumed absence of sharp absorption bands in quartz. In case, for example, the true maximum of the energy reflected from a given crystal fell within one of the quartz bands, it was shifted by this absorption toward longer or shorter wave-lengths as the case may be, the wave-length of the resulting maximum lying outside of the quartz band itself. Examining the curves from this point of view it seems possible, that, in the light of these new absorption bands, the center of the true reflection maximum of TlBr might lie at a wave-length slightly longer than 117μ. That of AgBr might in reality lie a bit below the reststrahlen value found at 112.7μ. The wave-length of the true reflection maximum of Hg_2Cl_2 might be longer than 98.8μ, and that of AgCl shorter than 81.5μ. In other words the absorption bands of the interferometer plates have caused in some cases an actual shift in the reststrahlen wave-lengths with respect to the wave-lengths of the true reflection maxima.[14] In using these reststrahlen later

[12] H. Rubens, Berl. Berichte, p. 513, 1913. Rubens found the transmission of a plate of quartz 7.393 mm $\parallel$ to be a few percent higher than that of a sample 7.256 mm $\perp$. This he attributed to dichroismus.

[13] See reference 12.

[14] It must be remembered that the reststrahlen wave-length has no theoretical meaning and cannot be calculated from the constants of the crystal in question. By definition, it is simply the maximum of the energy left over after the original radiation has suffered some form

In order to determine the absorption of quartz, Rubens had approximately the same amount of quartz in his apparatus as he did when he determined their wave-lengths. Therefore, the actual energy maxima used in curves *F* and *G*, agreed with the previously measured wave-lengths, and so, no measurement was made on quartz within a region of strong absorption. If the above assumptions are true, it is then obvious that by using the same rest-strahlen plates, with no quartz however in the apparatus, the bands of quartz might easily (have) appeared in curves *F* and *G*.

Comparison of fused, and crystalline quartz

For use as a filter or window material, fused quartz seems to be better suited than crystalline quartz. Because the absorption of the former is so very much stronger than that of the latter, a thinner and less expensive piece of the fused quartz may be used to produce the desired effects. Then too, since the bands are so much broader and less intense in the fused condition, any undesirable effects of these bands, such as the shifting of the true wave-lengths of the maxima or minima of the substance which is being measured, will be greatly reduced if fused quartz is used.

Upon comparing the two sets of curves the following correlations seem possible. If the absorptions in fused quartz arise from sources similar to those in crystal quartz, that is, from groups of crystallized SiO_2 molecules, then the effect of the actually fused material has been to broaden the bands and to shift them towards longer wave-lengths. The broad 86μ (fused) band may be the same as the sharp 77μ (cryst.) band, and the 118μ (fused) band the same as the 105μ (cryst.) band. The band indicated at $\lambda > 132\mu$ (fused) would be the 122μ (cryst.) band. Then, the supposed band at $\lambda > 132\mu$ (cryst.) would point to the presence of still a second unobserved band in fused quartz. These possibilities will be tested experimentally at a later date.

Regarding the Raman effect of quartz

Unfortunately, no very definite comparison can be made between these new absorption bands of crystalline quartz and the Raman effect. Czerny has shown that bands occur both in the Raman effect and in the infrared at 38μ and 78μ. At 48μ however a band occurs in the Raman effect which does not occur in the infrared. A comparison has also been made previously between the two sets of values at shorter wave-lengths. Since in the Raman effect the regions where $\lambda > 80\mu$ lie so close to the exciting line, the long wave-length values which have been obtained are extremely uncertain. Krishnan[15] mentions Raman lines at 94μ and 118μ, while Gross and Romanova[16] give $\lambda = 99\mu$, 120μ and 140μ. It is at once seen that none of these agree with the new infrared bands, possibly due simply to the inaccuracy in the R. E. values.

Gross and Romanova also investigated the Raman effect of fused quartz, and found that out to 47μ the indicated absorptions were practically the same

of selective reflection. It depends entirely upon the distribution of the energy in the source, the selective reflection and any selective absorption which may take place.

[15] K. S. Krishnan, Nature **122,** 477 (1928).

[16] E. Gross and M. Romanova, Zeits. f. Physik **55,** 744 (1929).

as in crystalline quartz. The chief difference in the two cases, was that the bands, which were sharp in crystalline quartz were broad and diffuse in fused quartz. Surprised at the occurrence of the same bands in the two cases they raised the question regarding the possible presence of crystalline structure in the fused condition. Since in x-ray photographs no crystal structure is found in fused quartz, they concluded that there could be no recrystallized centers in their samples.[17] This conclusion, which is rather doubtful, together with the fact that the same Raman lines were present in both forms of quartz, led them to conclude further regarding crystalline quartz, ". . . , that, in this region (8μ to 70μ) no frequency characteristic of the crystal lattice (outer vibration) is present". This second conclusion is surely wrong. They were unable to determine the exact locations, due to the weak dispersion of their apparatus, of any bands beyond 47μ in fused quartz. They state that the 78μ band probably does not occur, and since this, according to infrared theory, is characteristic of the quartz lattice, they hold its failing as evidence for their above mentioned conclusions. From the absorption curves however one sees clearly that the 78μ band should not occur in fused quartz at 78μ, but rather at 86μ. Too, the other long wave-length bands would not be expected to occur at the same wave-lengths as the corresponding ones in crystalline quartz.

It must be remarked however, that also in the infrared the absorption curves of fused and crystalline quartz are very similar below about 45μ, especially, of course, in regions where "inner vibrations" are present.

Sulphur, Fig. 5

The dotted curve *A*, which is taken from the work of Taylor and Rideal, shows, in a plate of rhombic sulphur 6.0 mm thick, bands at 7.76μ, 10.73μ,

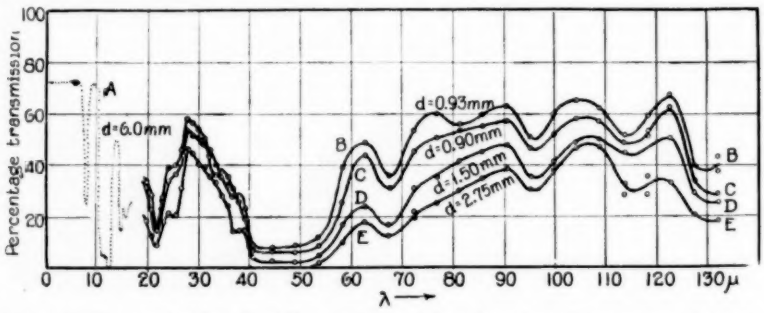


Fig. 5. Transmission of sulphur. Curve *A* is taken from Taylor and Rideal. All the curves shown here are for rhombic sulphur.

11.9μ and 14.6μ. Since in the present apparatus plates of sulphur 23 mm by 8 mm were required, considerable trouble was experienced in obtaining suitable samples. Thanks to the kindness of Professor Johnsen of the Mineralog-

[17] These recrystallized groups in order to be detected by the Debye method, would have to be relatively large, i.e., of the same size as finely powdered quartz crystals. Recrystallization centers, however, many times smaller than such powder grains could easily produce the infrared absorption bands characteristic of a large crystal, and still escape detection by x-ray analyses.

ical Institute, and of Dr. von Deines of the Physical Chemistry Institute, of the University of Berlin, crystals were obtained which were large enough to yield plates of the required size. Curves *B* (0.93 mm), *C*(0.90 mm), *D*(1.50 mm) and *E*(2.75 mm) show the transmission of four samples cut from rhombic sulphur. Of these, only sample *B* was cut parallel to the base of the crystal. The other three were each perpendicular to the base, the more exact crystallographic orientation being uncertain. It must be remarked here that sulphur is extremely brittle, and when only slightly heated often shatters completely. In the grinding and polishing of the plates, this was the chief difficulty encountered. The 2.75 mm plate *E* was cut from a natural crystal, the others coming from artificially grown samples.

Upon seeing that sulphur has such sharp and intense absorption bands throughout the infrared, one immediately asks two questions. Why does sulphur, an element, absorb in the infrared? Where does it get its electric moment? From x-ray analyses it has been shown that sulphur exists in the crystal form as S_{16}. Taylor and Rideal believe from their work that this is in reality $(S_2)_8$, and that each of the S_2's has an induced electric moment equivalent to about 0.7 times the moment it would have if one electron had actually been transferred from one *S* to the other. The binding between the two *S*'s is assumed to be of a pseudoheteropolar nature. The question however of exactly how and why a non-metallic element such as sulphur should have such strong infrared absorptions is still a puzzling one.

From the curves given here, one sees bands at 21.5μ, 25.0μ, 37.5μ, 40μ to 50μ, 67μ, 96μ, 114.5μ and $\lambda > 130\mu$. The region from 40μ to 50μ, when investigated with a finer grating, will undoubtedly show a complex structure. It is interesting to note that by the Raman effect[18] bands are found at 21.28μ, 23.06μ, 41.3μ, 46.25μ, 65.7μ, and 118.2μ. The agreement between the two sets of values is incomplete, the Raman effect failing to show the presence of the band beyond 130μ, and the ones at 96μ and 37.5μ. Since the infrared absorption was the same in natural and synthetic crystals, it seems improbable that these extra bands are due to impurities. This is however not impossible. Choosing as fundamentals his Raman effect bands at 21.3μ, 41.3μ and 65.6μ Krishnamurti has published an energy scheme which explains the bands found by Taylor and Rideal as overtones. It seems improbable, however, that the final and correct scheme should so entirely overlook these equally strong infrared absorptions which lie at longer wave-lengths.

It will be noticed that curve *B*, which is the only curve measured on a plate cut parallel to the base, differs from the other three in two places. At 81μ a weak band is present and the 37.5μ band occurs in this sample at 33μ. These differences were carefully checked and are real. It will also be noticed that although sample *B* was actually thicker than sample *C*, its percentage transmission throughout was always a few percent higher than that of *C*. This is analogous to the cases cited above for crystalline quartz and is probably here also an effect of dichroismus. Assumed differences in the degree of polish of the two crystals could also cause irregulaties in the percentages of trans-

[18] P. Krishnamurti, Indian J. of Phys. **5**, 105 (1930).

mission. This was here, however, very probably not the source of the effect, and in the case of quartz, certainly not.

Paraffin and mica, Fig. 6

The dotted curve *A'* shows the transmission of a sheet of 68°–72°C melting point paraffin[19] 0.6 mm thick. It shows that in thicknesses of approximately 1 mm paraffin would absorb rather thoroughly the radiations shorter than 18μ. Curves *A* and *B* were measured upon plates of paraffin 2.05 mm and 3.7 mm thick. From 20μ to 75μ the transparency increases uniformly, after which, absorption bands occur at 82.5μ, 106μ and 127.4μ.

Since paraffin may be had in so many various forms chemically, i.e., in forms varying widely in their melting points and the lengths of the chains of atoms, a test was made to see what effect these variations have on the transmission. From Kahlbaum samples were obtained having melting points at 68°–72°C and 42°–44°C, and from these two plates having exactly the same thicknesses, 2.2 mm, were cut. The results are shown in the curves marked

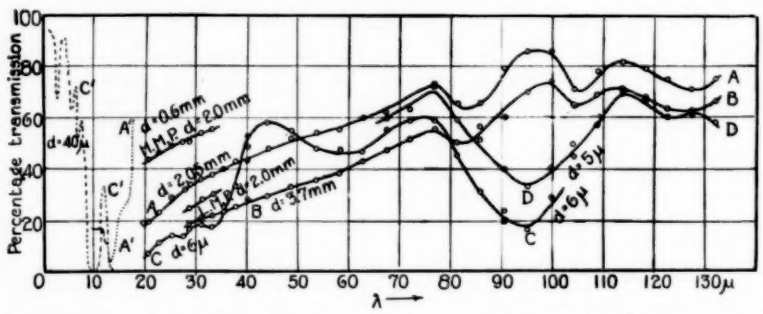


Fig. 6. Transmission of paraffin and mica. *A'*, *A*, *B*, *H.M.P.*, and *L.M.P.* are transmission curves of paraffin, *C'*, *C* and *D*, mica.

H.M.P. and L.M.P., the two forms varying approximately 100 percent in their transmissions. It is evident that curves *A* and *B* were made from L.M.P. paraffin. It will be interesting to see if the absorption bands which occur at long wave-lengths in L.M.P. paraffin are the same in H.M.P. paraffin. Frau Kellner measured a 2 mm thickness of H.M.P. paraffin from 20μ to 38μ, the results falling exactly with my (H.M.P.) curve.

Mica is chemically also a very poorly defined substance. Previous investigations in the near infrared have shown that the various micas differ somewhat in their absorption. Curve *C'* gives the transmission of a 40μ sheet of Muscovite mica as measured by Coblentz, and shows absorptions at 2.8μ, 5.5μ, 6.8μ, 10μ and $\lambda > 13\mu$. The samples used in the present work were of unknown origin and composition. It can only be said in this regard, that, in relatively thick sheets the mica had somewhat of a yellowish brown color while the thin samples used here were perfectly clear and colorless. The results are shown in curve *C* (6μ thick) and curve *D* (5μ thick). The samples used were split from thicker sheets, and their thicknesses measured roughly with a pocket spectroscope. It will be noted that the absorption is of a differ-

[19] L. Kellner, geb. Sperling, Zeits. f. Physik **56**, 215 (1929).

ent order of magnitude from that met with in the compounds discussed above. The absorption of the 5μ and 6μ films is more nearly of the same order of magnitude as that of NaCl or KCl. Even in such thicknesses the near infrared is almost completely absorbed, and bands occur at 27μ, 33μ, broad regions around 59μ and 95μ, and at 123μ. The values usually found in the literature are for thicker samples and often show irregularities due to the effects of interference between radiation reflected at the front and back surfaces. Here, owing to the thinness of the films, such effects are impossible.

Mica is particularly unsuited for use in the infrared. Paraffin, on the other hand, is very suitable for removing the near infrared. Below 80μ it is exceptionally good as a window material.

Miscellaneous, Fig. 7

Curve *A* shows the transmission of a film of celluloid of approximately 1μ thickness, formed by allowing one drop of a solution consisting of nitrocellulose and camphor dissolved in amylacetate and acetone to spread itself

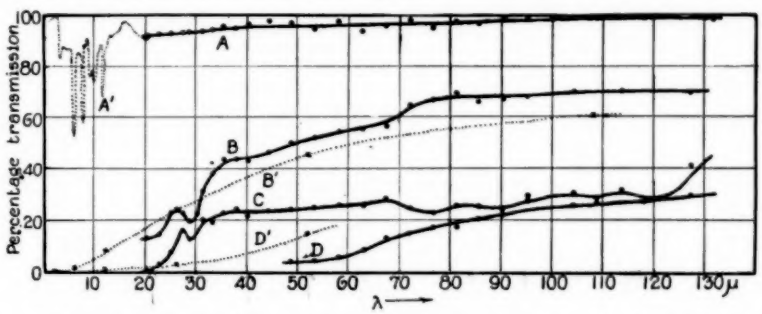


Fig. 7. Transmission of miscellaneous substances. *A'* and *A*, thin films of celluloid. *B'*, *B*, *D'* and *D*, black paper. *C*, a 40μ thickness of cellophane.

out upon a water surface. This curve simply corroborates what had already been stated.[20] Since these films can play such a large role in the infrared it is important to have an actual measured transmission curve covering the entire range of wave-lengths. The transmission is nowhere 100 percent, the losses being probably due chiefly to reflection and not to absorption. Curve *A'* taken from the work of Fräulein Tolksdorf shows the unsuitability of such a film for use in the near infrared.

Curve *B* is from a piece of black tissue paper which was slightly translucent for light from a 100 watt lamp. Beyond 35μ the transparency increases uniformly until 80μ is reached, and from there on to 135μ is practically horizontal. Below 35μ however, the entire style of the curve is different, showing at 28.7μ a rather sharp band. This band is probably due to the cellulose molecules themselves as it is exactly the same band as the cellophane band shown below it.

Curve *C* shows the measurements on a piece of cellophane, 0.04 mm thick, such as that used for wrapping candy boxes. From 20μ to 35μ the curve is qualitatively the same as curve *B*, the same band at 28.7μ occurring in both.

[20] M. Czerny, Zeits. f. Physik **34**, 227 (1925).

It is possible that the rest of the curve should simply be a straight line, but I rather believe that weak absorptions are present as shown at 75μ, 90μ and 121μ.

Curve D, which is for a piece of black paper, 0.10 mm thick, used for wrapping photographic plates, is almost identical qualitatively with the corresponding part of curve B.

The dotted curves B' and D' are from Rubens and Wood and represent measurements upon approximately similar samples of paper. The agreement with curves B and D is good.

Blackening substances, Fig. 8

Curve A, is from a film of camphor soot which was deposited upon a 1μ film of celluloid. Through this thickness of soot, the filament of a 100 watt lamp showed faintly with a deep red color. Beyond 30μ the transmission is

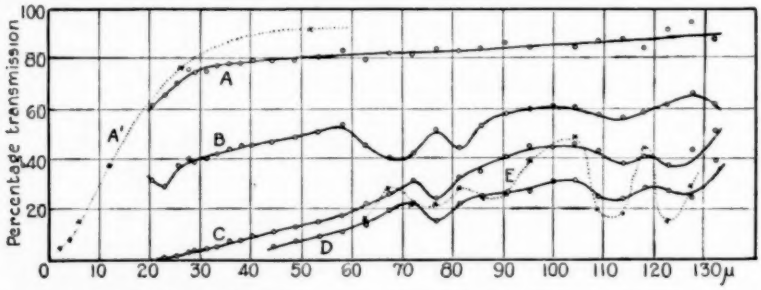


Fig. 8. Transmission of blackening substances. A' and A, soot films. B, Pfund's bismuth black. C, D and E, Rubens and Hoffman's soot-waterglass preparation.

uniform and high. Such a film is therefore very suitable for removing the visible and the extremely short infrared radiations. As a coating for receiving elements however, it is out of the question. Curve A', from Rubens and Wood, shows the transmission of a similar film made from candle soot. The extension of A would probably follow A' very closely.

Pfund[21] showed that if bismuth be evaporated in a vacuum of about 0.25 mm of mercury, a film is deposited upon the walls of the container which closely resembles black velvet in its blackness. Films of this bismuth-black, thick enough to be opaque to visible light, were found by Pfund to have transmissions under 1 percent, as far out as 13μ, and were therefore recommended by him for coating receivers for use in the near infrared. Curve B shows the transmission of such a film in the far infrared. It was deposited upon a 1μ film of celluloid, which during the depositing process had to be in contact with a plate of brass. (When this was not the case, no deposit was obtained.) Absorption bands exist at 22.5μ, 69μ, 81μ, 113μ and $\lambda > 130\mu$.

Curves C and D give the results for two films of a mixture of lampblack and waterglass approximately 0.07 mm and 0.08 mm thick. This mixture was prepared according to a method given by Rubens and Hoffmann,[22] and then painted upon thin celluloid films. These authors found that such a film

[21] A. H. Pfund, Rev. Scient. Instr. **1**, 397 (1930).

[22] H. Rubens and K. Hoffmann, Berl. Berichte p. 424, 1922.

became transparent first around 120μ. In fact they stated, that, ". . . a 50μ absorbing layer of lampblack and water glass out to wave-lengths of 120μ showed no real decrease (from 100 percent) of the absorption power, . . . ". However, the conditions under which they measured were quite different in one respect from those used in this present work. As a carrier for their mixture they used a film of mica 36μ thick, which according to them showed definitely interference effects. To obtain the effect due to the mica alone, the transmission for the various wave-lengths was measured before and after the mixture was painted upon it. In view of these interference effects which were present, it is questionable as to whether the values found "before" can simply be subtracted from those found "after" in order to obtain the transmission of the mixture itself. With one surface blackened the interference effects most probably disappeared, and the transmission of 36μ of mica was surely throughout most of this region, nearly zero. My samples of their mixture, approximately 70μ and 80μ in thicknesses, showed around 30 percent transmission beyond 70μ. Absorption bands were found at 76.8μ, 112μ and 125.0μ.

A rather interesting effect is shown by the dotted curve *E*. The sample shown by curve *D* was prepared, and after drying for 24 hours gave the curve *E*. Three days later an attempt to check this curve failed badly. On the next day still a third curve was obtained from the same sample. A new film, the one shown in curve *C*, was prepared and compared with the old one. The curves disagreed. After the second film was approximately a week old, the curves *C* and *D* were simultaneously measured. The agreement between these two showed that during the drying and hardening process, certain chemical changes had taken place, a final state having been reached after 4 or 5 days. It is well known that upon drying Na_2SiO_3 changes its condition considerably.

From the curves of this figure, one must conclude that up to the present we know of no ideal substance for coating receivers for the far infrared. Such a substance should show no sharp absorption bands, should absorb the far infrared entirely, and if possible should have a high percentage transmission for the near infrared. Czerny suggests a mixture of powdered silicates such as mica, glass, quartz and sodium silicate. Where one of the substances failed to absorb, some of the others probably would.

Conclusion

In conclusion I wish to state that in this paper no attempt has been made to present a complete and final investigation of the absorption of such substances as quartz and sulphur. To do this, special attention would have to be paid to such questions as polarization, the exact crystallographic orientation of the crystal, impurities, temperature effects and losses by reflection. The purpose of this paper is to present a general view of the transmission of the substances most commonly used in the spectroscopy of the far infrared.

I wish to take this opportunity to express my thanks to Dr. M. Czerny of the University of Berlin for the use of his apparatus and for his kindly interest in this work.

DIRECT MEASUREMENT OF MEAN LIVES OF ATOMIC STATES

By H. D. Koenig and A. Ellett
State University of Iowa

(Received December 9, 1931)

Abstract

A method is developed for measuring the mean lives of atomic states by means of the thermal motion of excited atoms in two forms of resonance lamps. The methods are used to determine the mean life of the cadmium 2^3P_1 state for which state the more accurate method gives $T=2.5\times10^{-6}$ sec. with a probable error which is difficult to determine but can hardly be in excess of 10 percent.

IT is obvious that measurements of the mean life of atomic states by means of the thermal motion of excited atoms in a resonance lamp are possible if the mean lives are long enough to make the distance which an excited atom

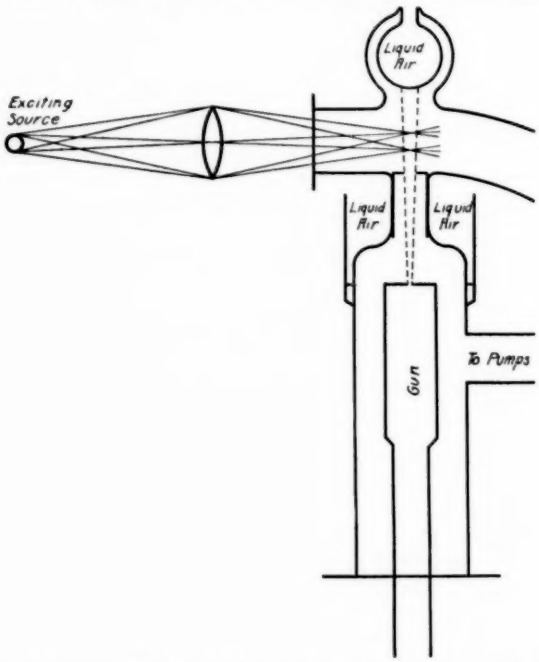


Fig. 1. Preliminary system. The only changes made throughout the entire procedure were to enter stops defining the exciting beam of radiation and to connect a second pump to the upper liquid air trap.

will move between excitation and emission observable by means of the camera with which the radiation is photographed.[1] Since the cadmium 2^3P_1

[1] Dunoyer attempted this in 1913. He shot a beam of sodium atoms through a beam of radiation of the sodium D lines. The mean lives of the excited states—the 2^2P states—were too short, however, to cause any observable motion of the excited atoms out of the excitation region. Dunoyer, Le Radium **10**, 400 (1913).

state was believed to have a mean life of sufficient length to permit such a measurement—Ellett obtaining the value 2.3×10^{-6} sec. by the method of depolarization of resonance radiation in weak magnetic fields—the following system was adopted to endeavor to measure the mean life of this state by a

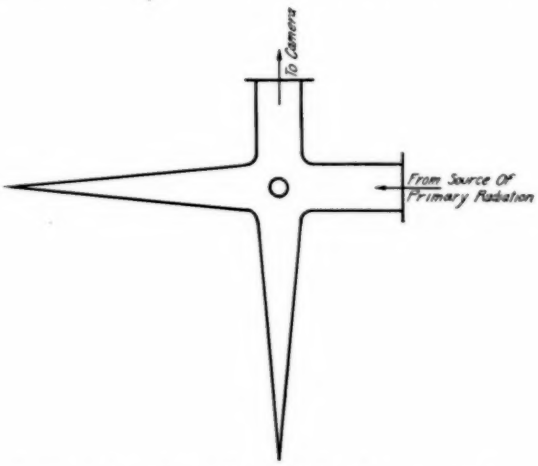


Fig. 2. Cross section of experimental chamber, as seen from above.

method depending upon the thermal motion of excited atoms. The apparatus, with the stops used to define the beam of exciting radiation removed, is shown in Figs. 1 and 2. The cadmium vapor effusing from a small hole in the top of a boiler—the gun of the diagram—was frozen out of the system with liquid air except for the small portion passing through the hole in an aluminum tube set in the constriction between the gun chamber and the upper chamber. The image of a cadmium discharge was focused upon the cadmium beam and the resonance radiation photographed along a line perpendicular to both cadmium beam and direction of exciting light ray.

Fig. 3. Print from plate taken in preliminary run.

The print shown in Fig. 3 is from a preliminary run taken with the apparatus, and is given because it illustrates the coarser characteristics of the method better than the plates taken later would. The arrow indicates the direction of the incident light ray, the broken line the direction of the cad-

mium beam. Collisions between the cadmium atoms in the beam and air molecules in the experimental chamber caused sufficient vapor pressure of cadmium to develop in the experimental chamber to define the path of the exciting light ray by resonance radiation from this vapor. Before the more accurate work was attempted, a second pump was added to the system to eliminate this effect. At the position where the cadmium beam passed through the exciting ray the radiation is noticeably more intense and extends about 1.5 mm above the upper edge of the exciting beam.

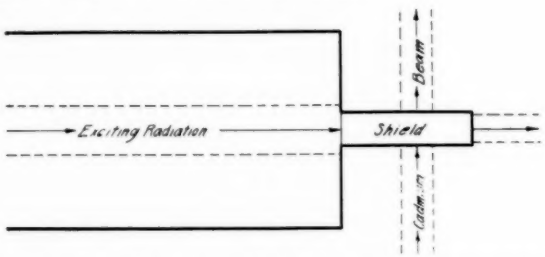


Fig. 4. Cross sections of slit system used for first method of determining the mean life.

To make use of such a system, however, it is necessary, first, to make the intensity of the exciting radiation break off sharply at the upper edge of the excitation region as we move along the beam; second, to obtain a system which will show a definite place to start from in making measurements on the films; third, to demonstrate that the effect is due to motion of excited atoms and not secondary resonance radiation. To meet these requirements the slit system of Fig. 4 was inserted into the experimental chamber of the apparatus. An improvement in the pumping system was made at the same time and the gas pressure in the experimental chamber reduced to less than 5×10^{-6} mm Hg.

Fig. 5. Print of plate taken for first method of determining mean life.

In this system the exciting radiation from the discharge[2] strikes a slit and then passes behind a shield which hides the path of the radiation from the camera. This shield was made so as to extend 0.5 mm beyond the edge of the excitation region in order to make adjustment easier and give complete assurance that the path of the exciting radiation was hidden from the camera. The cadmium beam passed through the exciting radiation behind this shield.

[2] The discharge was of the type described by Ellett in connection with his earlier work on the depolarization of resonance radiation in magnetic fields. Ellett, Jour. Opt. Soc. Am. **10**, 427 (1925).

In Fig. 5 the upright lines mark the edges of the cadmium beam and the arrow indicates the direction of motion of the atoms. The lines drawn out to the side from these indicate the top of the shield. These marks were drawn on the plate. The radiation photographed is observed to start at the marks indicating the upper edge of the shield and to extend upward, decreasing in intensity quite rapidly as we go up along the beam.

This radiation was now entirely due to motion of the excited atoms since the entire excitation region was hidden from the camera and the failure of any radiation to appear below the shield eliminates the possibility of the effect being due to secondary resonance radiation.

Microphotometer curves of plates obtained by this method were taken. Assuming that the blackening on these plates was proportional to the intensity of radiation and that all the atoms were moving with the arithmetic mean velocity, the intensity of radiation outside the excitation region should be $I=I_0e^{-t/\tau}$ where t is the time the atom has been out of the excitation region and I_0 the intensity at the edge of the excitation region. Then if x is the distance from the edge of the excitation region along the direction of motion of the cadmium atoms, $t=x/v$ and by our assumption v is the arithmetic mean velocity.

Then

$$I = I_0e^{-x/v\tau}.$$

And the mean velocity of the atoms may be calculated from the temperature of the boiler which was measured by a copper-constantin thermocouple inserted in the base of the boiler for the number of atoms having a velocity between v and $v+dv$ effusing from the boiler is proportional to

$$e^{-mv^2/2kT}v^3dv$$

and if the probability that one of these is excited is proportional to $1/v$ the number of excited atoms having velocities between v and $v+dv$ is proportional to

$$e^{-mv^2/2kT}v^2dv.$$

So for the excited atoms

$$v(\text{mean value}) = \frac{\int_0^\infty e^{-mv^2/2kT}v^3dv}{\int_0^\infty e^{-mv^2/2kT}v^2dv}.$$

The assumption that the probability of excitation is proportional to $1/v$ is justified by the fact that the time any of these atoms remains in the excitation region is short and proportional to $1/v$. By fitting curves of this form to the microphotometer curve, the mean life of cadmium 2^3P_1 was calculated to be $\tau=1.3\times10^{-6}$ sec. The assumptions used, however, are not sufficiently accurate to regard this as more than a measurement of the order of magnitude, probably correct to within a factor of 2.

In this regard some justification must be given for giving this value to the 2^3P_1 state. Except in the case of very high excitation, 2^3P_1 and 2^1P_1 are the only states excited to any noticeable degree. It is a matter of experience with resonance radiation excited by a discharge of the type used that after the discharge has operated for some time its ability to excite the 2^1P_1 state decreases probably due to the formation of a very thin layer of cadmium glass in the quartz section from which the exciting radiation is taken. Since the radiation obtained in the cadmium beam was of very low intensity, the exposure time for the films was about four hours, 2^3P_1 and 2^1P_1 are the only states which could be excited to any degree, and 2^3P_1 is chosen because of the characteristics of the source of exciting radiation.

The difficulties met here in connection with the faults of the plate can be overcome if we can obtain a system in which the position of the maximum intensity is a function of the mean life τ and variables depending upon the geometry of the apparatus and the temperature of the cadmium atoms. In the system just discussed the greatest intensity was at $x=0$ regardless of either the mean life or the gun temperature. To obtain a set up of this type, let us consider a system in which the number of excited atoms having a velocity between v and $v+dv$ and position between x and $x+dx$ is N where x is measured along the atomic beam from an arbitrary plane perpendicular to the beam. Then since there is no appreciable secondary resonance radiation we may assume that the actual system consists of a group of independent systems each having a definite velocity—or range of velocities—and the contribution of the group to the intensity of radiation will be the sum of the contributions of the individual systems.

Then $(N/\tau)dtdx$ is the number radiating between $x=x$ and $x=x+dx$, $t=t$ and $t=t+dt$; and if $f(x)dxdt$ is the probability that an atom will be excited between x and $x+dx$, t and $t+dt$, the number excited within these limits will be $f(x)(N_g/v)dxdt$. Here N_g is the number of atoms leaving the gun and entering the beam with a velocity between v and $v+dv$ in one second, and in any experiment we can carry out $N_g/v \gg N$.

Now if we follow a group of atoms along the beam and disregard forced radiation due to light falling upon excited atoms $(dN/dt)dx = f(x)(N_g/v)dx - (N/\tau)dx$, or considering the forced radiation

$$(dN/dt)dx = f(x)(N_g/v)dx - (N/\tau)dx - Bf(x)Ndx$$

where B is the B_{2-1} coefficient of Einstein's radiation laws.

But $Bf(x)$ is small compared to $1/\tau$ so

$$\frac{\partial N}{\partial x}v + \frac{\partial N}{\partial t} = f(x)\frac{N_g}{v} - \frac{N}{\tau}$$

and N is independent of t so $\partial N/\partial t = 0$

$$\frac{dN}{dx} + \frac{N}{v\tau} = \frac{N_g}{v^2}f(x), \quad N = e^{-x/v\tau}\int_0^x e^{x/v\tau}\frac{N_g}{v^2}f(x)dx$$

if $N=0$ at $x=0$.

That is, if the excitation region starts at $x=0$.

Let $f(x)$ be a continuously decreasing function finite at $x=0$ and 0 for $x \geqq a$.

Then since $N=0$ at $x=0$ and is decreasing and greater than 0 at $x=a$, there must be a maximum value of N between $x=0$ and $x=a$. The function N must be decreasing at $x=a$ since N must reduce to $kf(x)$, (k a constant) for small values of $v\tau$ or values of x large compared to $v\tau$.

For the system actually used $f(x)=A(1-ax)$, $a=1.15$. The slit system used is shown in Fig. 6.

Two straight edges were set so as to form an acute angle and make the portion of the cadmium beam exposed to the exciting radiation triangular. Then if we observe the region between two planes perpendicular to the direc-

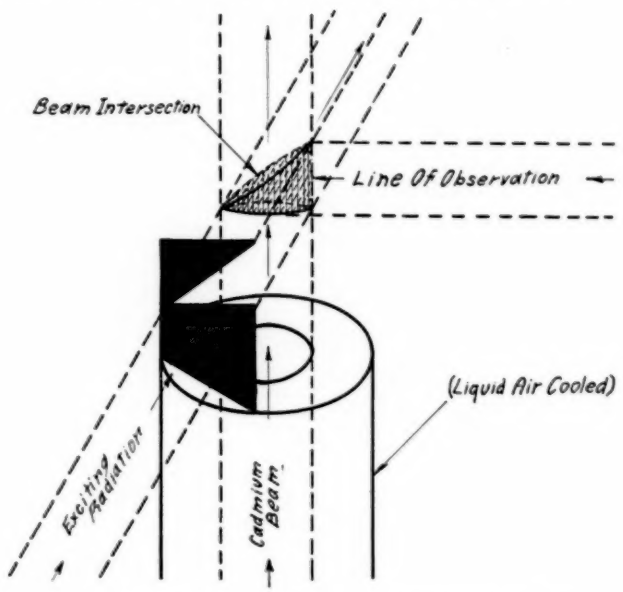


Fig. 6. Cross sections of slit system for second method.

tion or the exciting radiation and near the mid part of the cadmium beam as seen by the camera, the number of atoms excited per second between $x=x$ and $x=x+dx$ is proportional to the depth—measured along the line of sight—of the portion of the cadmium beam subject to exciting radiation. This is then proportional to $(b-cx)$ where b is the greatest width of the portion of the cadmium beam subject to excitation and c is the slope of the upper straight edge referred to the line of sight.

After the exciting radiation has passed through the cadmium beam, it struck an upright. This upright—a small wire nail—was heavily blackened to prevent it from scattering too much light. Then the portion of this nail which was illuminated by the exciting radiation was photographed with the resonance radiation and used to define the lower edge of the excitation region, the point chosen as our zero value of x.

Fig. 7 shows the print of a plate taken with this apparatus. The upright on the left is the slit. The upright on the right is the exciting ray defining nail.

Fig. 7. Print of plate taken for second method of determining mean life.

Two arrows drawn on the plate show the direction of the cadmium beam. The radiation is seen above and between the arrows.

A microphotometer curve was taken for this plate and is shown as Fig. 8.

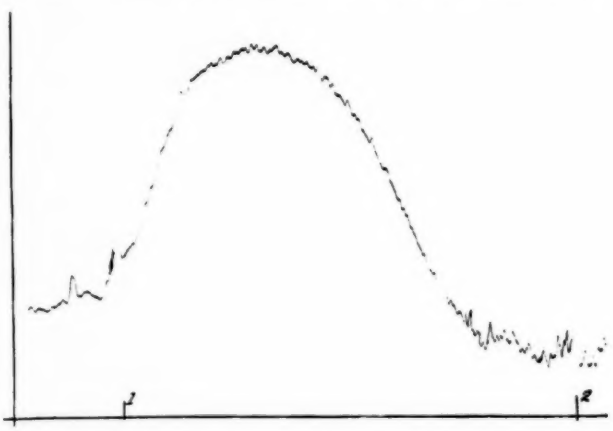


Fig. 8. Photometer curve taken off same plate as Fig. 7. Instrument run in direction of increasing x—direction of arrows on Fig. 7.

Now for the system represented here

$$I = N/\tau = A \int_0^\infty e^{-h m v^2} v^3 e^{-x/v\tau} \int_0^x e^{-x/v\tau} \frac{(1 - ax)}{v^2} dx dv$$

where $h = 1/2kT$, $a = c/b = 1.15$, $T = 690°\mathrm{K}$.

$$N/\tau = I = A \int_0^\infty e^{-h m v^2/\tau}(- v\tau + 1.15 x v\tau - 1.15 v^2\tau^2) v dv$$

$$+ A \int_0^\infty e^{-h m v^2/\tau} e^{-x/v\tau}(v\tau + 1.15 v^2\tau^2) v dv.$$

For a maximum

$$dI/dx = 0 = A \left[1.15 \int_0^\infty e^{-h m v^2} v^2 dv - \int_0^\infty e^{-h m v^2} e^{-x/v\tau}(v/\tau)(1 + 1.15 v\tau) dv \right]$$

The first integral is known and the second was evaluated by assuming it equal to

$$\int_a^\infty e^{-h m v^2}(v/\tau)(1 + 1.15 v\tau)dv$$

for large values of v and integrating mechanically over smaller values of v.

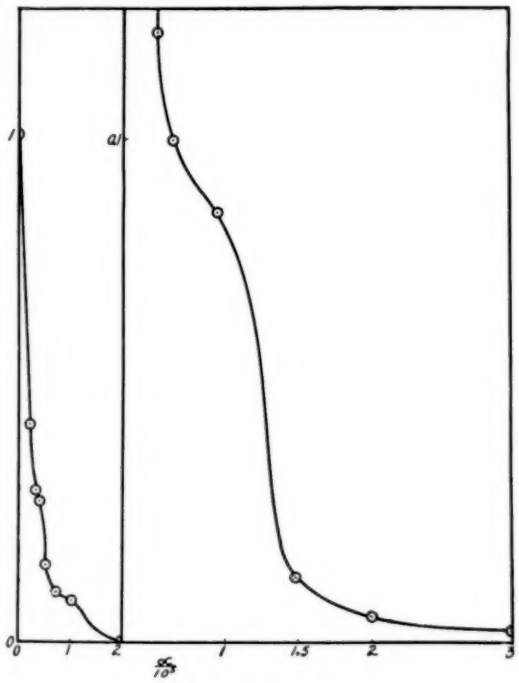


Fig. 9.

$$\frac{\int_0^{10^5} e^{-h m v^2} e^{-\alpha/v} v^2 dv}{\int_0^\infty e^{-h m v^2} v^2 dv}$$

plotted against $\alpha_0/10^5$.

For the mechanical integration the integral

$$\int_0^{10^5} e^{-h m v^2} e^{-\alpha/v} v dv, \quad \alpha = x/\tau$$

and the integral

$$\int_0^{10^5} e^{-h m v^2} e^{-\alpha/v} v^2 dv$$

were integrated by plotting curves for $\alpha = 10^4$, 2×10^4, 3×10^4, 4×10^4, 5×10^4, 7×10^4, 10^5, 1.5×10^5, 2×10^5, 3×10^5 and integrating with a planimeter. The value of the integral was then plotted as a function of α—Figs. 9 and 10—and

the value of X, 0.46 cm—was taken from the photometer curves—Fig. 8. This made τ the only variable of the equation for the position of the maximum.

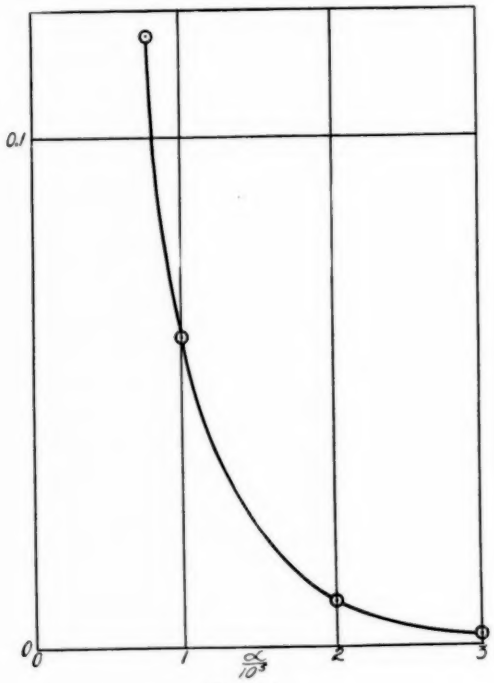


Fig. 10.

$$\frac{\int_0^{10^5} e^{-hmv^2} e^{-\alpha/v} v^2 dv}{\int_0^{\infty} e^{-hmv^2} v^2 dv}$$

plotted against $\alpha/10^5$.

The mean life was then determined by graphing

$$Y = \tau \int_0^{\infty} e^{-h\,mv^2} v^2 dv - \int_0^{\infty} \frac{e^{-h\,mv^2} e^{-x/v\tau}}{1.15} v dv$$

$$- \tau \int_0^{\infty} e^{-h\,mv^2} e^{-x/v\tau} v^2 dv$$

as a function of τ. The zero value of this function gave the desired value of τ, being $\tau = 2.5 \times 10^{-6}$ seconds.

In conclusion the authors wish to acknowledge an obligation to Professor Tate of the University of Minnesota who placed the microphotometer used in this work at their disposal.

INELASTIC AND ELASTIC ELECTRON SCATTERING IN ARGON

By A. L. Hughes* and J. H. McMillen
Washington University, St. Louis

(Received January 7, 1932)

Abstract

The energy distribution and the angular distribution of electrons scattered by argon atoms were investigated both for elastically and inelastically scattered electrons. *Elastic scattering* was investigated over the angular range between 10° and 170°, and for the energy range between 50 and 550 volts. The scattering curves fell off steeply with increasing angle. In the case of the 50 and 100 volt electrons, however, maxima were found at 100° and 90° respectively. The 400 and 550 volt curves when plotted as a function of sin $(\theta/2)/\lambda$ (where θ is the angle and λ the de Broglie wavelength) were superposable, in agreement with Mott's theory. The number of elastically scattered electrons, integrated over all angles, depends on the colliding energy in much the same way as does the difference between the total electron absorption coefficient and the ionization efficiency. The energy distributions of electrons scattered *inelastically* in argon at 10° were measured for 50, 100, and 200 volt electrons. It was found that as the energy of the colliding electrons increased the probability of the larger energy losses became relatively greater than the probability of the smaller energy losses. Angular distribution curves (5° to 35°) for such losses (11.6 to 34.0 volts) were steeper the smaller the magnitude of the loss. For any one loss, the steepness increased with the speed of the colliding electron. Electrons which have lost more than half the energy left over after ionization are called *ejected* electrons. To each ejected electron there corresponds a *colliding* electron, and the sum of their energies amounts to the original energy before collision less the energy of ionization. The angular distributions of slow ejected electrons having various amounts of energy (1 to 8 volts) were studied for different collision energies (50 to 200 volts). Such distribution curves showed a general distribution of electrons over all angles with, in many cases, strong maxima at large angles (90° to 160°). The positions of the maxima depended in a definite way upon the energy of collision and the energy of the ejected electron.

THE study of the scattering of electrons by gaseous atoms has received a new impetus in recent years (*a*) because of the remarkable experimental results of Davisson and Germer, of Thomson, and of Rupp, and (*b*) because some aspects of the problem have been interpreted successfully from a new theoretical standpoint, that of wave mechanics.

Mott[1] calculated the angular distribution of electrons scattered by gas atoms by means of wave mechanics. He arrives at the result that the "scattering coefficient", that is, the number of electrons scattered in a direction θ from that of the original beam through unit solid angle, from unit length of path of the original beam, per single electron in this beam, per single atom per unit volume, is

$$\alpha(\theta, v) = \left[\frac{e^2}{2mv^2}(Z - F)\ \mathrm{cosec}^2\frac{\theta}{2}\right]^2 \qquad (1)$$

* This research was assisted by a grant made to the senior author from the Science Research Fund of Washington University.

[1] N. F. Mott, Proc. Roy. Soc. **A127,** 658 (1930).

when e, m, and v, are the charge, mass, and velocity of the electron, Z the atomic number and F the atomic structure factor. Mott pointed out that there is a striking similarity between the wave mechanical theory of the scattering of electrons and the ordinary theory of scattering of x-rays, the formula for the latter case being

$$\alpha'(\theta) = \left[\frac{e^2}{mc^2} Z \operatorname{cosec}^2 \frac{\theta}{2}\right]^2. \tag{2}$$

Important experimental investigations of electron scattering have been carried out recently by Arnot[2] and by Bullard and Massey.[3] They find that, in the region between 30° and 120°, the scattering coefficient shows maxima and minima which are especially well marked for the lower velocities. Mott's theory predicts a rapid decrease in the scattering coefficient with increasing angle, but fails to predict the presence of the maxima and minima. This is due to the omission of certain effects which are negligible only at high velocities. These effects—electron exchange, polarization of the atom by the electron, and distortion of the incident electron wave—, which should be considered in a more complete theory, modify the theoretical scattering curve especially at low velocities, and qualitatively account for the presence of the maxima and minima.[4] At sufficiently high velocities, Mott's theory is in satisfactory agreement with experiment.

In addition to the elastically scattered electrons, there are many varieties of inelastically scattered electrons which have lost different amounts of energy as a result of exciting or ionizing the atom. A complete experimental exploration of the field of excitation losses would call for a study of the scattering of many distinct groups of electrons, each group corresponding to a particular energy loss which is determined by the energy level to which the excited atom is raised. Spectroscopy shows that there are many of these energy levels and it is probable that there is a characteristic angular distribution for the electrons which have been effective in raising the atom to the particular energy level considered.[5] We should also consider the scattering of electrons which have ionized atoms. In the absence of evidence to the contrary, we may assume that the *colliding* electron may lose any amount of energy *greater* than that just necessary for ionization. The electron *ejected* from the atom in the ionization process will, if conservation principles apply and if the positive ion has but one energy value, have an amount of energy just equal to that lost by the colliding electron over and above the amount necessary for ionization A complete analysis of the problem would call for a study of the angular distribution of both the colliding and ejected electrons over the full range of possible energies. As it is impossible to distinguish be-

[2] F. L. Arnot, Proc. Roy. Soc. **A129,** 361 (1930); **A130,** 655 (1931); **A133,** 615 (1931).

[3] E. C. Bullard and H. S. W. Massey, Proc. Roy. Soc. **A130,** 579 (1931); **A133,** 637 (1931).

[4] F. L. Arnot, Proc. Roy. Soc. **A133,** 615 (1931); E. C. Bullard and H. S. W. Massey, Proc. Roy. Soc. **A133,** 637 (1931).

[5] For an interesting theoretical treatment of this problem, see H. S. W. Massey and C. B. O. Mohr, Proc. Roy. Soc. **A132,** 605 (1931).

tween the identity of two electrons appearing after the ionization of an atom, it is convenient for the present to call the faster electron the colliding electron and the slower one the ejected electron.[6]

The study of the scattering of electrons is closely related to two other fields, (1) the absorption of electrons by a gas, and (2) the ionization efficiencies of electrons of various energies. The integral of all electrons, scattered elastically and inelastically, taken from a certain small angle up to 180°, measures the total number of electrons diverted from the beam and therefore the absorption of the electrons by the gas. The total number of ionizing and ejected electrons, integrated from 0° to 180°, should be equal to twice the number of ions produced. (The direct measurements of the absorption and of the ionization efficiencies, of course, give far more accurate results than could be obtained by integration of the scattering curves.)

Experimental Method

The apparatus used is shown in Fig. 1. It consists of two parts, a collision chamber B and an analysing chamber C, the latter sorting out electrons of

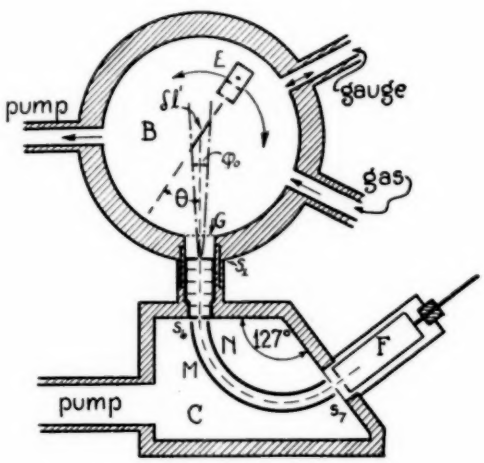


Fig. 1. Apparatus.

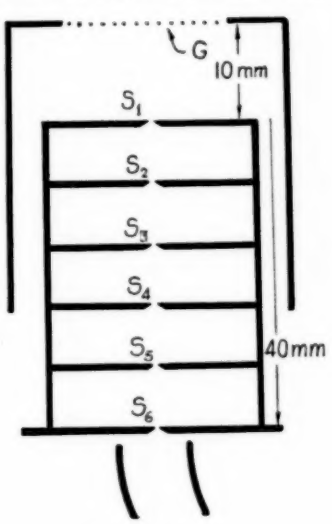


Fig. 2. Slit system.

different energies. The collision chamber is made of brass with three gas inlets, one connected to a diffusion pump, a second to a McLeod gauge and the third through a fine capillary to a reservoir containing pure argon. This constant renewing of the gas ensured a high degree of purity in the collision chamber. The pressure could be held constant at any desired value, the pressures actually used in the experiments to be described being between 0.0015

[6] If the ionization of a single atom by an electron could be studied, it would be interesting to find whether or not the angle of scattering of the ejected electron is uniquely determined by the angle of scattering of the colliding electron. It does not have to be uniquely determined, for the positive ion may share the energy and momentum lost by the colliding electron with the ejected electron in an infinite number of ways.

mm and 0.0030 mm. By means of a ground glass joint (not shown) the electron gun E could be rotated so that the main beam of electrons could be set at any angle θ to the beam of scattered electrons passing into the analyser. The analyser is one in which electrons of different energies are sorted out by a radial electrostatic field. As the analyser has been described at length in other papers, no further discussion will be given here.[7] The collision chamber and electron gun are substantially the same as those described in fuller detail by McMillen,[8] except for the fact that in the present apparatus the gun could be rotated as far as the 170° position. The analyser was separated from the collision chamber by a slit system (see Fig. 2) designed to offer considerable resistance to gas flow (because of the number of slits), so as to allow a very low pressure to be maintained in the analyser which was connected to a high speed diffusion pump. Since S_1 and S_6 were slightly smaller than the other slits, they defined the electron beam entering the analyser. The chief innovation in this apparatus is a grid in front of the entrance slits of the analyser which allows one, by means of a suitable field, to accelerate electrons up to the slits. The grid G is made up of fine platinum wires 0.001 inches in diameter, set about 0.01 inches apart. The wires run perpendicularly to the lengths of the slits and not parallel to them as implied in the diagram. The analysers used by McMillen and Van Atta could not be used for low velocity electrons. For some reason, not fully understood, it was not possible to get measurable effects when electrons of energy less than about 40 volts were directed on to the slits. It appeared as though the slits became sufficiently charged in some way to refuse admission to slowly moving electrons. Possibly electrons get entangled on an invisible film of grease or other insulating material over the slit surfaces and so abnormally large spurious contact potentials are formed. The obvious way to minimize this effect is to outgas the whole system at a high temperature, but this was impossible on account of the way the apparatus was assembled.[9] Consequently to study the distribution of slow electrons it is necessary to speed them up to an arbitrarily chosen speed. This was effected by applying the proper accelerating potential between the grid G and the slit S_1. The usual procedure in studying the energy distributions of a group of electrons was to set the potential between the deflecting plates M and N to deviate electrons of 100 volts energy from the lower entrance slit S_6 to the exit slit S_7. Then electrons of velocity V_0 approaching the grid perpendicularly in the collision chamber would have to be accelerated by a potential $(100-V_0)$ volts in order to register in the Faraday cylinder. The experimental variable was therefore the accelerating voltage $100-V_0$, from which the original energy of the electrons in the collision cham-

[7] A. L. Hughes and J. H. McMillen, Phys. Rev. **34,** 291 (1929); A. L. Hughes and V. Rojansky, Phys. Rev. **34,** 284 (1929); J. H. McMillen, Phys. Rev. **36,** 1034 (1930); L. C. Van Atta, Phys. Rev. **38,** 876 (1931); E. Rudberg, Proc. Roy. Soc. **A129,** 628; (1930); **A130,** 182 (1930).

[8] J. H. McMillen, Phys. Rev. **36,** 1034 (1930).

[9] See J. D. Whitney, Phys. Rev. **34,** 923 (1929); D. C. Rose, Canad. Journ. of Research **3,** 174 (1930).

ber could be inferred. The presence of the field between the grid and the first entrance slit has a focusing effect (when $V_0 < 100$), the result of which is to exaggerate the number of slower electrons present. It can be shown that the effect of an accelerating voltage V is to change the angle ϕ_0 between two electron paths into ϕ_1 where

$$\phi_1 = \phi_0 \frac{1}{(1 + V/V_0)^{1/2}}$$

and where V_0 is the initial energy of the electrons in volts (Fig. 3). (It is assumed that the angles are small enough to allow us to use the angles themselves in place of their sines and tangents.) The grid and slit system are drawn

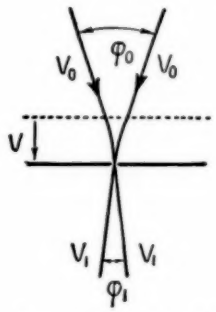


Fig. 3. Schematic representation of focusing effect.

approximately to scale in Fig. 2. From the dimensions given it is possible to show that ϕ_1 (width of S_6 divided by 40 mm) is 40 minutes. This angle ϕ_1 is fixed and consequently the angle ϕ_0 depends on the values of V and V_0. Since ϕ_0 in Fig. 1 determines $\delta l'$, the length of path of the original electron beam from which scattering into the analyser takes place, it is evident that in order to get comparable quantities we must *multiply* the experimentally measured electron current into the Faraday cylinder, N_θ'', by the factor

$$\frac{1}{(1 + V/V_0)^{1/2}}. \tag{3}$$

This gives us a corrected value N_θ'. This same length of path, $\delta l'$, also depends on the angle of scattering θ, and increases with $\sin\theta$ (Fig. 1). Hence to reduce the scattering to some constant path length, δl, we must multiply by $\sin\theta$. Thus the true value, N_θ, the number of electrons scattered from a constant length of path, δl, is

$$N_\theta = N_\theta' \sin\theta = N_\theta'' \frac{\sin\theta}{(1 + V/V_0)^{1/2}}. \tag{4}$$

The scattering coefficient is defined by the following equation

$$N_\theta = \alpha(V, \theta), N_0 \cdot \delta l \cdot \delta\omega \cdot n$$

where N_θ is the number of electrons scattered through an angle θ, within a solid angle $\delta\omega$, from a length δl of a beam of electrons containing N_0 electrons, when there are n atoms per unit volume. The scattering coefficient, $\alpha(V,\theta)$, defined in this way, is a function of the angle θ, and V, the energy of the electrons. It depends also on the nature of the scattering atom. Except when there is a statement to the contrary, the ordinates in the figures showing the experimental results graphically are proportional to this scattering coefficient $\alpha(V,\theta)$.

Elastic Scattering

The scattering coefficient for electrons scattered without loss of energy has been investigated over a range of electron energies between 50 and 550 volts. The results are shown in Fig. 4. For small angles, *i.e.*, for angles less

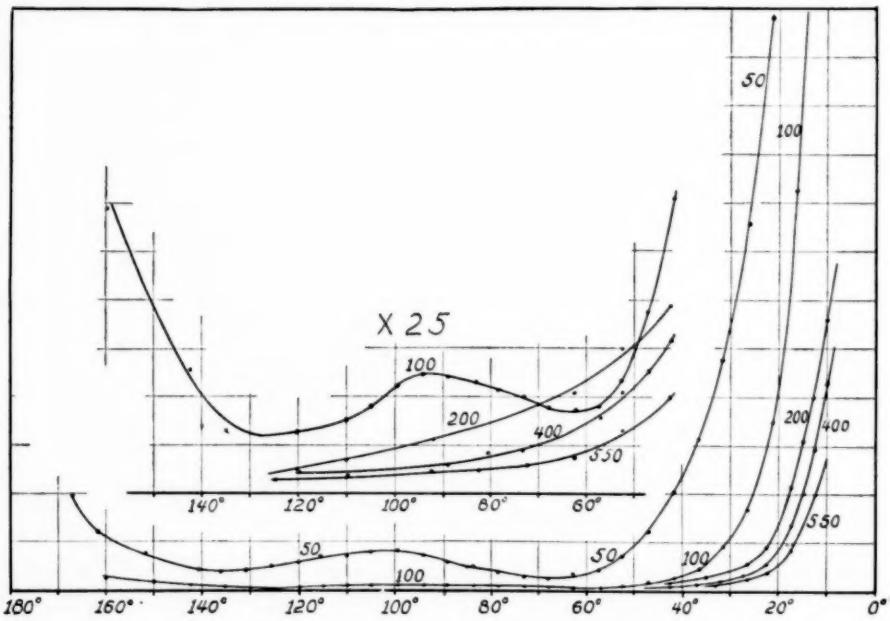


Fig. 4. Elastic scattering in argon.

than about 40°, the scattering coefficient decreases rapidly with increasing angle, the decrease being faster the greater the energy of the electrons. For the 50 and 100 volt energies we find a definite maximum in each curve in the neighborhood of 100°. The positions of these maxima are in satisfactory accord with the positions found by Arnot.[10] His range did not extend beyond 120°. In our experiments the range extended to 170°. It is interesting to note the marked increase in the scattering between 140° and 170°, a result predicted by Holtsmark's theory.[11]

Mott[12] derived the formula given in Eq. (1) for the scattering of electrons

[10] F. L. Arnot, Proc. Roy. Soc. **A133**, 615 (1931).

[11] See Fig. 7 in the paper by E. C. Bullard and H. S. W. Massey, Proc. Roy. Soc. **A133**, 637 (1931).

[12] N. F. Mott, Proc. Roy. Soc. **A127**, 658 (1930).

by individual atoms using the method of wave mechanics. Since the electron wave-length $\lambda = h/mv$, the expression can be changed into

$$\alpha(\theta, v) = \left[\frac{e^2m}{2h^2}(Z - F)\frac{1}{\mu^2}\right]^2 \tag{5}$$

where $\mu = \sin(\theta/2)/\lambda$, a quantity upon which x-ray scattering has been shown to depend. The scattering coefficient should therefore have the same value for the same μ, no matter what the velocity of the electron may be, provided we take such a value of θ to give the same μ. Our results are given in Fig. 5. It is clear that the higher the energy of the colliding electrons the closer do

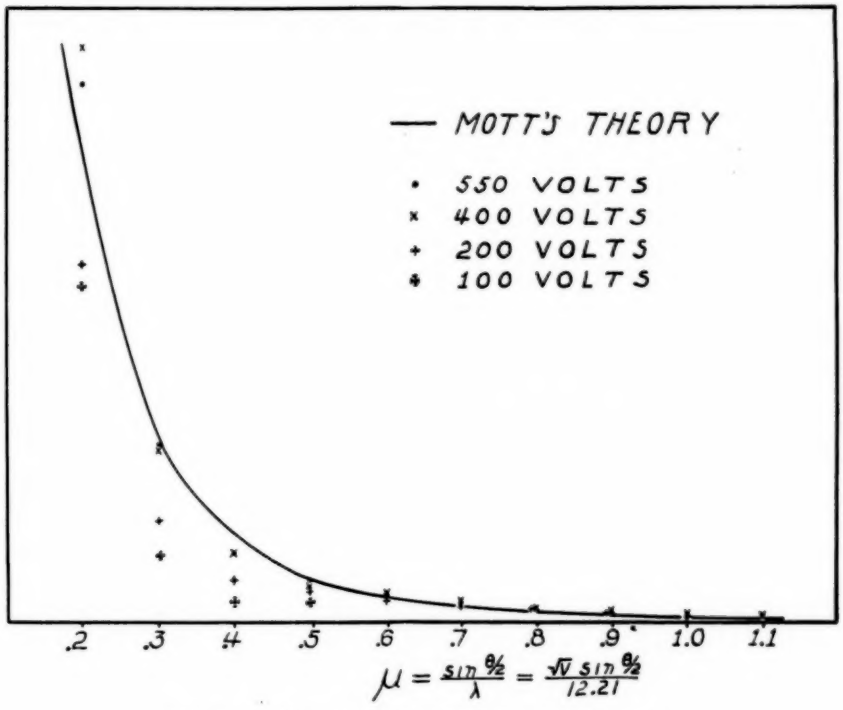


Fig. 5. Elastic scattering plotted as a function of μ.
Mott's theoretical curve is plotted on half scale.

the experimental points come to the theoretical curve. Mott pointed out that his formula should agree better with experiment the higher the energy of collision. These experiments then indicate that, for scattering of electrons by argon atoms, Mott's theory gives an accurate description of the facts when the electron energy exceeds about 400 volts.[13,14]

[13] The deviation of the experimental results from the theoretical curve for 400 and 550 volt electrons, at $\mu = 0.2$, is explained by the fact that this corresponds to a very small angle where the experimental errors are considerable.

[14] The F values used in plotting the curves were computed by James using the Hartree field for the argon atom. We wish to thank Mr. N. F. Mott for communicating these unpublished results to us.

We attempted to determine the value of the scattering coefficient in absolute units. This is a matter of considerable difficulty, for it involves assumptions as to the geometry of the apparatus and especially as to the *effective* dimensions of the very narrow slits S_1, S_6, and S_7. The final result, however for the 500 volt electrons, gave a value only 50 percent less than that calculated by Mott, which is perhaps the best we can expect in view of the fact that the apparatus was certainly not designed to give good absolute determinations.

While the work was in progress the very interesting and comprehensive results of Arnot and of Bullard and Massey were published. We therefore temporarily postponed our studies of elastic scattering and turned to the hitherto unexplored territory of inelastic scattering.

Inelastic Collisions

The amounts of energy lost by electrons in inelastic collisions differ according to the type of collision. The least amount of energy which can be lost is that corresponding to the first radiation potential of argon, viz., 11.6 volts;

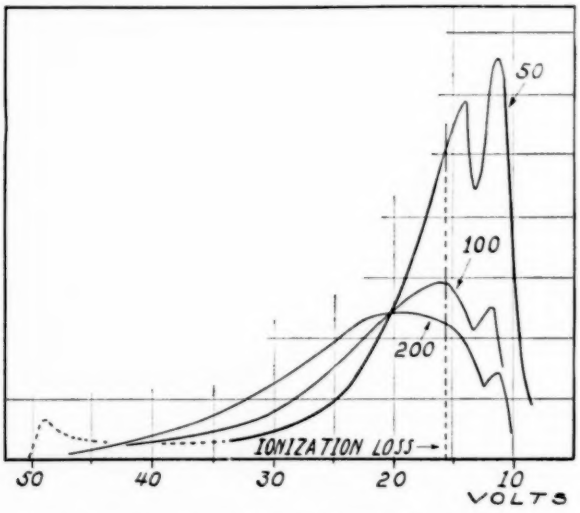


Fig. 6. Energy losses in argon. Angle of scattering 10°.

then follow a series of greater losses corresponding to excitation of the atom to higher levels. A "continuous spectrum" of energy losses is to be expected when ionization of the atom is effected, for presumably the energy left over after ionization may be divided in an infinite variety of ways between the original electron and ejected electron.

Typical distributions of inelastic energy losses for electrons colliding with different speeds, all scattered at 10^0, are shown in Fig. 6. The 11.6 volt loss stands out distinctly, but the resolution of the apparatus was not sufficiently good to isolate other excitation losses. The curves are arbitrarily fitted together at an energy loss of 20 volts (i.e., 4.5 volts above the ionization potential). The curves have been corrected for the focusing effect by means of Eq.

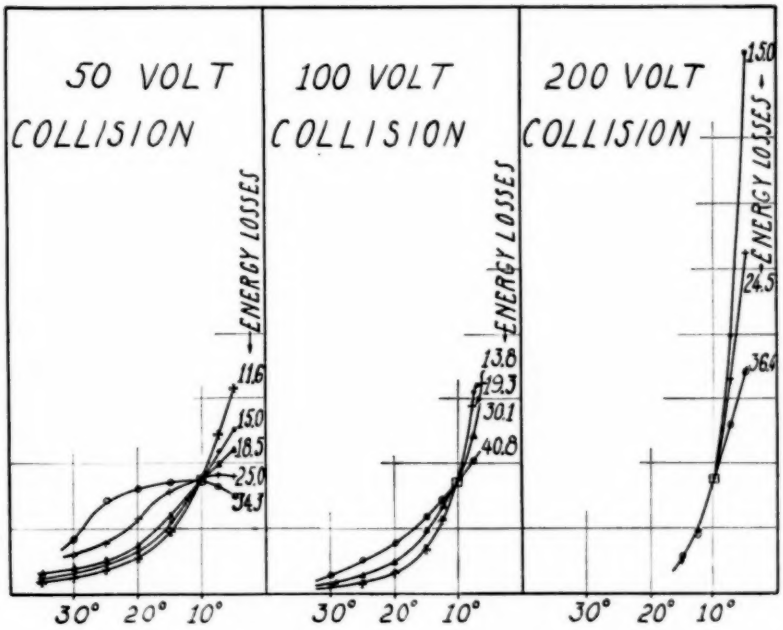


Fig. 7. Angular distributions of various energy losses for 50, 100 and 200 volt collisions.

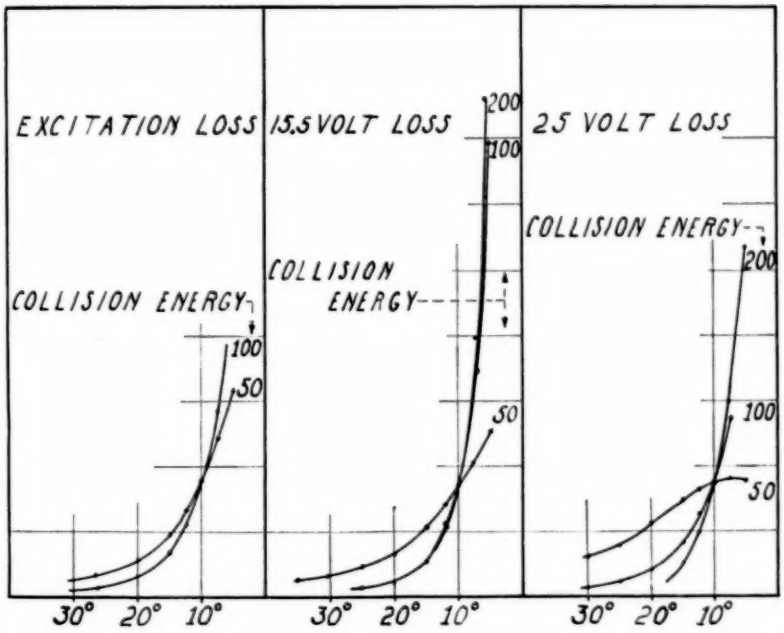


Fig. 8. Dependence of angular distributions on the collision energy for different energy losses.

(3), and for slight absorption of the electrons in the gas. Both corrections were small. It is clear that as the energy of the colliding electron increases, (1) the number of collisions producing the 11.6 volt excitation diminishes relatively, and (2) the number of collisions in which the colliding electron ionizes the atom and loses a considerable amount of energy to the ejected electron over and above that necessary for ionization, increases relatively to the total number of collisions.

The angle distributions of various losses were measured and the results are shown in Figs. 7 and 8. In Fig. 7 we compare the angular distribution curves for various energy losses, first when the electrons in the primary beam have 50 volt energy, then when they have 100 volt energy, and finally when they have 200 volt energy. In Fig. 8 we have picked out a particular energy loss and shown how the corresponding angular distribution curve depends on the

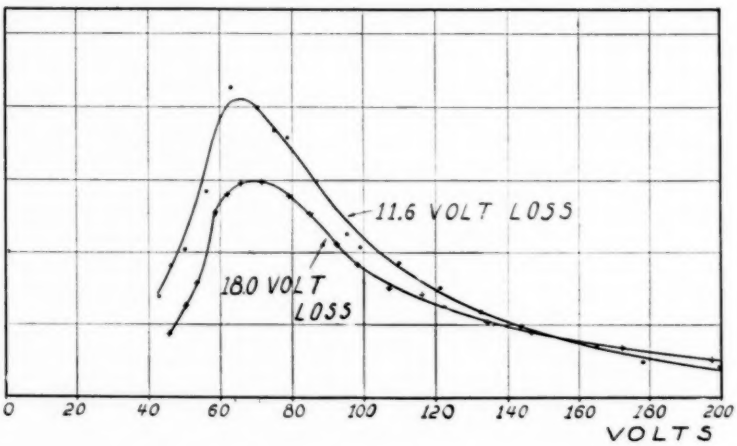


Fig. 9. The 11.6 and 18 volt energy losses, at 10°, plotted as a function of the collision energy.

energy of the electron before collision. The results may be summarized as follows: The *lower* the energy loss, the *steeper* is the angular distribution curve. (This is particularly true for the lowest collision energy, viz., 50 volts.) For any given energy loss, the angular distribution curves are *steeper* the greater the energy of impact.

The scattering coefficient for the 11.6 volt loss and for the 18.0 volt loss is plotted as a function of the energy of impact in Fig. 9, the angle of scattering in all cases being 10°. It is evident that primary electrons having 63 volts energy result in the maximum number of 11.6 volt energy losses, while primary electrons having 70 volts energy give the greatest number of 18.0 volt energy losses. (This statement covers only the case where the electrons are scattered through 10°; further systematic experiments are required to determine whether or not it is true for other angles of scattering as well.)

Attention is called to the fact that in our experiments the 11.6 volt loss has been found over a wide range of collision energies (Fig. 9) and over a wide

range of angles (Figs. 7 and 8). Bullard and Massey[15] were unable to find this loss in their experiments and quoted several others (Dymond, Michels, Oppenheimer) to the effect that this particular loss should not occur since the corresponding energy level is metastable, and so could be appreciably excited only by voltages very close to the excitation potential itself. However, we have found strong excitation over a wide range of voltages and associated with a fairly wide range of scattering angles. Well marked peaks corresponding to the 11.6 volt loss were also found by Van Atta[16] for zero scattering angle, and for a range of colliding energies from 100 to 300 volts. Van Atta took especial care to identify the peak and concluded that the energy loss was 11.53 ± 0.05 volts, which is in close agreement with the value 11.57 volts obtained from spectroscopic data. This agreement makes it difficult to maintain that the identification of the peak in our experiments and in those of Van Atta's is open to question. We should like to have traced the 11.6 volt

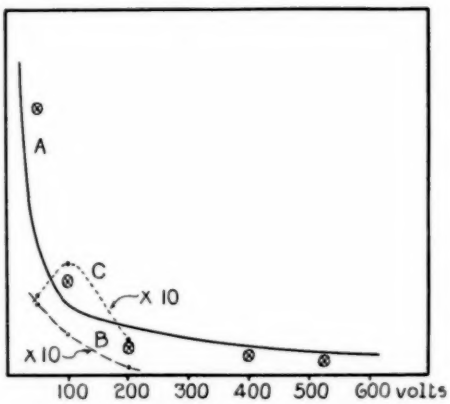


Fig. 9a. Absorption coefficients as computed from scattering curves. *A*, total absorption coefficient less ionization efficiency (Normand; Smith); *B*, "excitation" absorption coefficient; *C*, "ionization" absorption coefficient; *crosses*, "elastic" absorption coefficient. ($\times 10$ means that the real ordinates of *B* and *C* have been increased tenfold.)

energy loss curve in Fig. 9 right down to a colliding energy of that amount, as it is conceivable that a strong maximum would be found with colliding electrons of energy a volt or two above 11.6 volts, but experimental difficulties made it impossible to explore this region.

By multiplying the ordinates of the elastic scattering curves (Fig. 4) by $\sin\theta$ and then integrating over all angles, we get a measure of the total number of electrons lost from the initial beam through elastic collisions. The results of such an integration are plotted as a function of the electron energy in Fig. 9*a*. The total absorption coefficient of electrons as ordinarily measured gives the number of electrons lost from the beam through elastic scattering and inelastic scattering, the latter being associated with excitation and ionization. Since the number of collisions associated with excitation losses is rela-

[15] E. C. Bullard and H. S. W. Massey, Proc. Roy. Soc. **A130,** 583 (1931).
[16] L. C. Van Atta, Phys. Rev. **38,** 876 (1931).

tively small, an absorption coefficient for elastic scattering could be obtained by subtracting from the total absorption coefficient as determined by Normand,[17] the total ionization efficiency as measured by Smith.[18] The result of such a subtraction is plotted on an arbitrary scale in Fig. 9a. The agreement with our results is fairly satisfactory, when the uncertainty involved in integrating over the very small and the very large angles is realized. A somewhat similar procedure allows us to calculate the total number of electrons associated with all the excitation losses and with all the ionization losses, leading to what may be called the "excitation absorption coefficient" and the "ionization absorption coefficient". The results are included in Fig. 9a. The ionization absorption coefficient so calculated has the same general shape as that determined by direct experiment. It is not possible to obtain accurate values of the various absorption coefficients in this way as considerable errors are possible in the graphical integration.

Ejected Electrons

An electron of energy V collides with an atom and loses energy amounting to V'. If $V' > V_+$, the ionization potential, then ionization takes place and the ejected electron goes off with energy $V' - V_+$. The energy left to the colliding electron is $V - V'$. (The colliding and ejected electrons have identical energies when they share, equally between them, the surplus left over after ionization, and this has a value $\frac{1}{2}(V - V_+)$.) We shall refer to the faster of the two electrons as the colliding electron and to the slower as the ejected electron. These are to be regarded as convenient labels, for we do not mean to insist that the electron going away from the atom with the greater speed after ionization is necessarily identical with the electron which hit the atom and caused the ionization. In a complete experimental description of the process of ionization we should need to include the angular distributions of all colliding and ejected electrons.[19]

In Fig. 6 we show all the possible energy losses after a 50 volt collision, resulting in electron energies after collision right down to zero. (Such low energies after collision are not shown for the 100 and the 200 volt collisions because we used a scale suitable for displaying certain other features.) We shall now consider the experimental distributions of the low velocity electrons appearing after various kinds of collisions. It will be remembered that N_θ'', the experimentally measured current of electrons entering the Faraday cylinder must be multiplied by the factor $1/(1+V/V_0)^{1/2}$ in order to correct for the focusing action of the accelerating field between G and S_1. This factor becomes very different from unity for ejected electrons of very small energies. As a concrete case we may state that in one set of experiments electrons were ac-

[17] C. E. Normand, Phys. Rev. **35**, 1217 (1930).

[18] P. T. Smith, Phys. Rev. **36**, 1293 (1930).

[19] The ideal method would be to study single encounters and to measure θ for the colliding electron and ϕ for the ejected electron, and to find out whether or not ϕ is determined uniquely when θ is given. There is possibly no unique value for ϕ, because for a given θ there may be many combinations of ϕ and of ψ the angle of recoil of the atom.

celerated up to a constant energy of 100 volts. Electrons of 100 volts energy would require no accelerating voltage V, while electrons of 1 volt energy would require an accelerating voltage $V=99$ volts. In the first case, no correction would be needed, i.e., $N_\theta' = N_\theta''$, whereas in the latter case the correction would be given by $N_\theta' = 1/10\ N_\theta''$. While we believe that the theory for the correction factor applies to the whole range, it is evident that if it is only approximate, the errors due to its use will be relatively greater the larger the ratio V/V_0. Thus it is possible that the numerical values for N_θ for ejected electrons of low energy are not strictly comparable with those for much faster electrons.[20]

In Fig. 10 we have plotted the distributions of energies of the slow electrons ejected by electrons of different speeds. The continuous lines represent the experimentally measured electron current to the Faraday cylinder, i.e.,

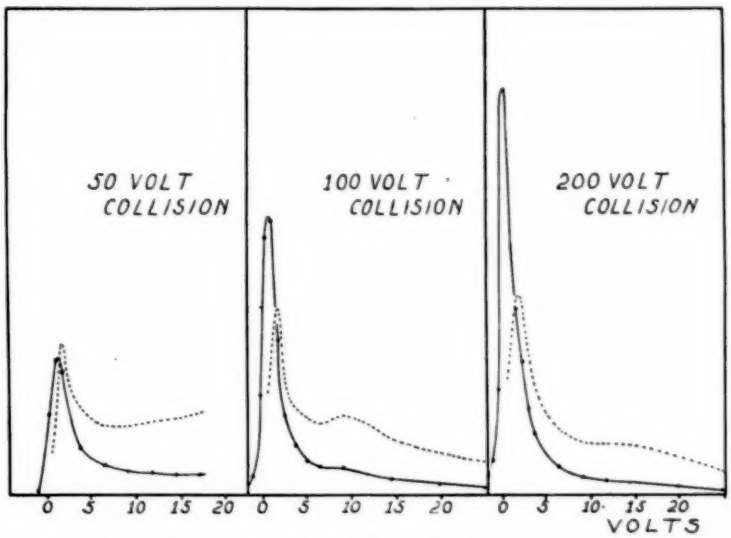


Fig. 10. Energy distribution (at 10°) for electrons ejected at 50, 100 and 200 volt collisions.

N_θ'', while the broken lines represent the values, N_θ', of the electron current after correction for focusing has been made. (The values of N_θ' have been increased by a scale factor of 10, otherwise the characteristics of the curve would hardly have been observable. The three pairs of curves are not to the same scale, however.) A survey of the curves shown in Fig. 10 shows that the slower electrons are the more plentiful. The peaks indicating the most probable energy of the ejected electrons occur between 0.5 and 1.0 volts. However, the resolution for slow electrons analysed by the method of introducing an

[20] It was possible to check the theory giving the factor $1/(1+V/V_0)^{1/2}$ over a moderate range by carrying out some experiments in which no accelerating field was applied between G and S_1. Since the analyser would not register electrons entering it if their energies were below about 40 volts, it was impossible to check the theory for slower electrons.

accelerating field between G and S_1 is poor, and it may well be that a more accurate investigation would show the peaks to come exactly at zero energy.[21]

To each colliding electron which has lost an amount of energy V', i.e., $V'-V_+$ over and above the energy necessary for ionization, there should correspond an electron ejected with energy $V'-V_+$. Hence the total number of each should be equal. Experiment gives us N_θ'' for each. As in Eq. (4) we get the number scattered per unit solid angle by multiplying by $\sin\theta$ and by the focusing factor $1/(1+V/V_0)^{1/2}$. To get the *total* number scattered in *all* directions we multiply by $\sin\theta$ again and integrate from 0° to 180°. On doing this we get pairs of curves analogous to those shown in Fig. 11. All that we can say is that the area under each curve is of the same order. One source of uncertainty lies in the fact that the smaller the angle, the larger is the number of colliding electrons scattered (see Fig. 7) and the measurements become increasingly inaccurate as we go to smaller angles, especially below 20°. Then

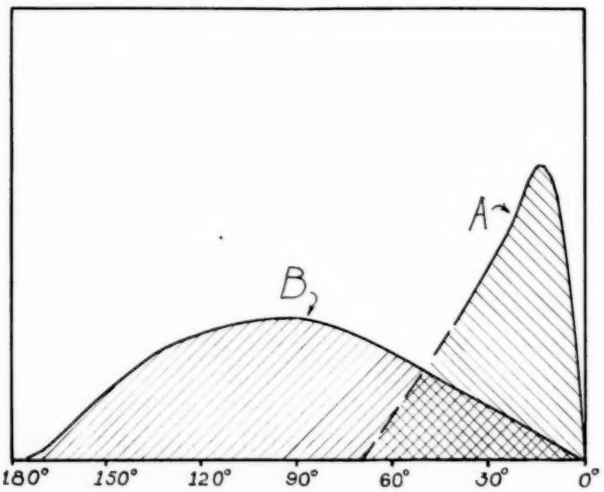


Fig. 11. A, Total number of electrons losing energy 3 volts more than energy required for ionization. B, Total number of ejected electrons having 3 volts energy. (The ordinates here measure the number of electrons scattered between the cones θ and $\theta+d\theta$.)

[21] F. L. Arnot, Proc. Roy. Soc. **A129,** 361 (1930); Proc. Camb. Phil. Soc. **27,** 73 (1931) has shown that when a beam of electrons is sent through a gas at a low pressure in a supposedly field-free space, a small field is set up between the inside of the beam and points at some distance from it, because the secondary electrons leave the beam with greater speeds than the positive ions. Such a field may distort the paths of electrons whose angular distribution we are investigating and this effect will clearly be larger the lower the energy of the scattered electrons. This effect should therefore be much more marked with our slowest ejected electrons than with the faster colliding electrons. Arnot found that the difference of potential set up between the inside of a beam and a point far outside it was proportional to the strength of the current in the beam and amounted to 2.0 volts for an electron current of 25 microamperes through mercury vapor at about 0.001 mm. As our currents were generally less than 0.5 microampere, it seems justifiable in a preliminary survey of inelastic scattering to assume that the effect discovered by Arnot could be neglected.

again it is not certain that we can assume that the focusing effect is completely taken care of by the factor $1/(1+V/V_0)^{1/2}$, when we apply it to a range of electron energy extending from, say, 100 volts down to 0.5 volt. However, the agreement is perhaps as good as can be expected under the circumstances.

It should be noted that the integral of the number of ejected electrons having energies from 0 to $\frac{1}{2}(V-V_+)$ volts, or the integral of colliding electrons having energies from $V-V_+$ down to $\frac{1}{2}(V-V_+)$ volts (integrated in each case over all solid angles) should give the ionization efficiency for elec-

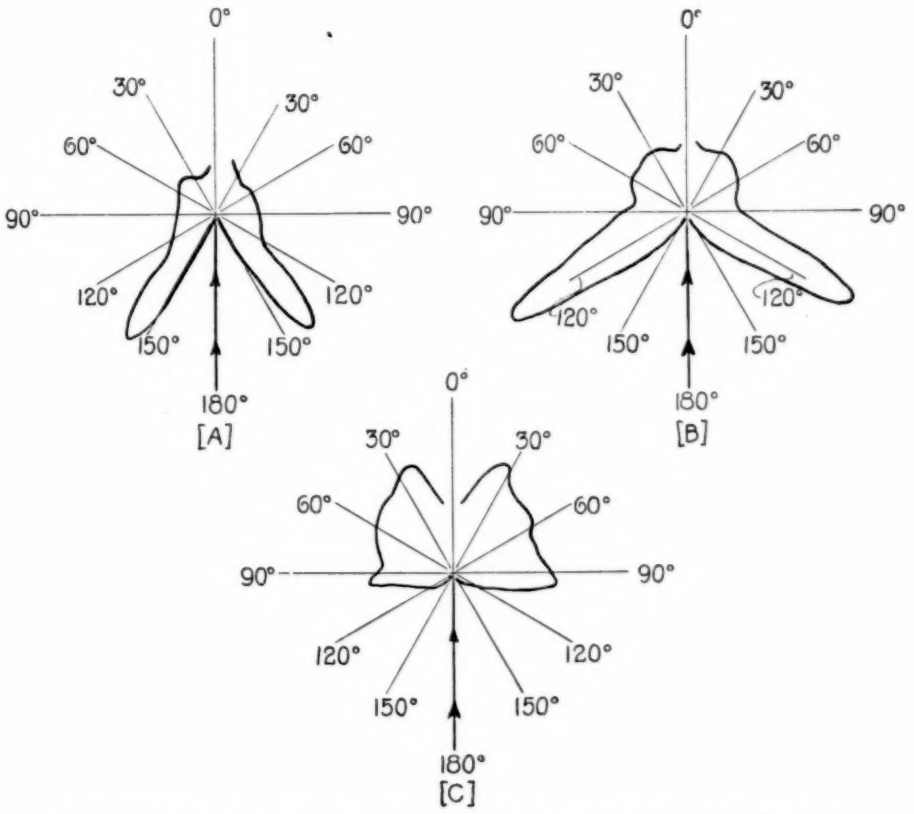


Fig. 12. Angular distribution of electrons ejected with 1 volt energy for (*A*) 50 volt collisions, (*B*) 100 volt collisions, and (*C*) 200 volt collisions.

trons colliding with energy V. More direct methods of course give the efficiencies far more exactly.

We now proceed to discuss the angular distributions of ejected electrons of energies 1, 3, 5.5, and 8 volts, when these are produced by primary electrons having energies 50, 100, and 200 volts. The results for the 1 volt ejected electrons are most strikingly displayed on polar diagrams (Fig. 12). The angular distributions are symmetrical on both sides of the main beam, as they should be. The results for all are plotted in Fig. 13. Before going into detail we may make the following general deductions.

(A). In striking contrast with the angular distributions of colliding electrons, in which the concentration is very high at small angles, i.e., in the for-

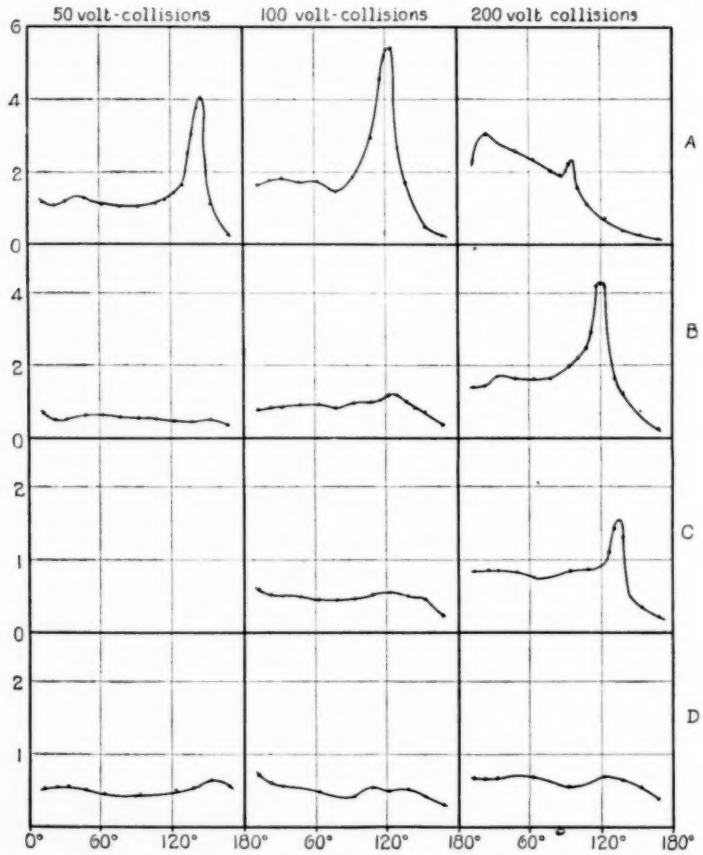


Fig. 13. Angular distribution of ejected electrons. *A*, 1 volt ejected electrons; *B*, 3 volt ejected electrons; *C*, 5.5 volt ejected electrons, *D*, 8 volt ejected electrons. (Note: vertical scale for *C* and *D* differs from that for *A* and *B*.)

ward direction (Fig. 7), the angular distributions of the ejected electrons are more evenly spread over a wide range of angles, with well-marked peaks at large angles in some cases.

A NEW DETERMINATION OF e/m FROM THE ZEEMAN EFFECT

By J. S. Campbell and W. V. Houston
California Institute of Technology

(Received December 19, 1931)

Abstract

Values of e/m have been determined from the Zeeman separations of the Cd line 6439 and the Zn line 6362. For these lines the g-factors can be accurately determined from the theory. A magnetic field of 7300 gauss was produced by an air core solenoid in which the variation of field strength over a length of 6 cm at the center was less than 0.1 percent. The field to current ratio, K, of this solenoid was determined in terms of the calculated ratios of three single layer standard solenoids. The field strength during an exposure was then given by the product of this constant and the current flowing. Evaporation of Zn and Cd in the short (6 cm) positive column of a helium discharge tube gave the desired lines. The Zeeman patterns were photographed with a Fabry-Perot interferometer. The result is

$$e/m = 1.7579 \pm 0.0025 \times 10^7 \text{ e.m.u. per gram.}$$

BECAUSE of the apparent discrepancy, emphasized by Birge,[1] between the values of the electronic specific charge obtained from experiments on cathode rays and those obtained from spectroscopic measurements, a precise determination of this constant has become of considerable importance.

Up to about 1927 the various determinations made with cathode rays, although rather widely scattered, seemed to be converging to a value near 1.768×10^7 e.m.u., and most reviews[2] of the subject ignored the value of 1.761 obtained by Babcock.[3] In 1927 a determination made by means of an interferometric measurement of the Rydberg constants of H and He^+ agreed so well with Babcock's result,[4] and a careful cathode-ray measurement by Wolf[5] showed such good agreement with the expected higher value, that Birge was led to suggest that the two types of measurement were really measuring different things. Subsequent theoretical work[6] has failed to show any justification for this difference on the basis of the present theory, so that it remained an open and important question as to what is the correct value of e/m.

This investigation was undertaken to make a precise measurement of the Zeeman effect under conditions which differed as much as possible in detail from those of Babcock. The result which has been obtained is

[1] R. T. Birge, Phys. Rev. Supplement **1**, 47 (1929).
[2] W. Gerlach, Handbuch der Physik **22**, 41–82 (1926).
[3] H. D. Babcock, Astrophys. J. **58**, 149 (1923); **69**, 43 (1929).
[4] W. V. Houston, Phys. Rev. **30**, 608 (1927).
[5] F. Wolf, Ann. d. Physik **83**, 849 (1927).
[6] See among others L. D. Huff, Phys. Rev. **38**, 501 (1931).

$$e/m = 1.7579 \pm 0.0025 \times 10^7 \text{ e.m.u. per gram}$$

which confirms the previous spectroscopic values, although it is a little lower than either of them.

Since this work was started, two papers have appeared which describe new cathode-ray measurements of the specific charge.[7] These results agree so well with the spectroscopic measurements that one is led to suspect that the "spectroscopic value" is correct, and that the discrepancy between the two methods was due to overestimation of the accuracy of the deflection methods.

General Requirements of a Zeeman Effect Determination

In order that a precise measurement of e/m can be made by means of the Zeeman effect it is necessary that the following three conditions be fulfilled.

(1). The magnetic field must be such that it can be accurately measured, reproduced, and kept constant during an exposure. It must also be uniform over a space large enough to contain the source of light. This was attained by the use of an air core solenoid.

(2). The spectroscopic measurements must be made with the desired precision. In the case of very simple patterns this can be done with a Fabry-Perot interferometer. The Zeeman pattern of a singlet line, when the source is viewed parallel to the magnetic lines of force, is a doublet, and so is ideal for use with this instrument.

(3). The Zeeman pattern of the lines used must be known with sufficient accuracy.

This, combined with the above requirement for the use of an interferometer, and the general requirement that the lines shall be sharp and as free as possible from hyper-fine structure, considerably restricts the selection of possible lines. Those finally selected were the $^1P - ^1D$ lines, Cd λ6439, and Zn λ6362. The fairly high atomic weight of these elements tended to keep the Doppler broadening within reasonable limits, even though the temperature of the source was fairly high. At no time was any trace of hyperfine structure observed, although of course the source was not particularly adapted to showing it.

For these singlet lines the Zeeman pattern can be determined with considerable accuracy.[8] It is affected slightly by the nearness of the ^3P and ^3D levels, so that the g-factor of each term is slightly more than one. The amount of this correction can be determined with more accuracy than is needed. For cadmium $g_D = 1.00049$, and $g_P = 1.00216$. For zinc $g_D = 1.00003$, and $g_P = 1.0002$. States assigned to other electron configurations might be expected to produce a slight additional change in these quantities, but since the nearest states are from configurations with a different parity and hence have no effect at all, and since the effect of the next states certainly would be very small, these can be neglected.

[7] C. T. Perry and E. L. Chaffee, Phys. Rev. **36**, 904 (1930); F. Kirchner, Phys. Zeits. **31**, 1074 (1930).

[8] W. V. Houston, Phys. Rev. **33**, 297 (1929).

When the g-factors of both the initial and final states are unity, the longitudinal effect shows a doublet. When, however, these factors differ slightly from this value, the various components of the line do not overlap exactly, and the pattern becomes very complex. Since this complexity is not resolved, the position of the center of gravity must be calculated in order to interpret the measurements. For each elementary component of the line, the displacement due to the magnetic field is

$$\Delta\nu_{ij} = aeH/4\pi mc \text{ where } a = m_i g - m_j' g'. \tag{1}$$

g and g' are the splitting factors for the initial and the final states respectively, while m_i' and m_j are the corresponding magnetic quantum numbers. For the whole complex line the expression for the displacement of the center of gravity is

$$\Delta\nu = \bar{a}eH/4\pi mc \text{ where } \bar{a} = \sum_{ij} I_{ij}(m_i g - m_j' g')/\sum_{ij} I_{ij}. \tag{2}$$

In this expression I_{ij} is the intensity of the transition $m_i - m_j'$. These can be determined from the usual formulae.[9] For the cadmium line used, $\bar{a} = 0.99966$, and for the zinc line $\bar{a} = 0.999945$.

The result is then that

$$e/m = 4\pi c\Delta\nu/\bar{a}H. \tag{3}$$

This equation gives e/m in absolute electromagnetic units of charge per gram. Here e is the velocity of light in a vacuum, H is the magnetic field in absolute gauss, and $\Delta\nu$ is the displacement of the center of gravity in cm^{-1} reduced to vacuum.

Apparatus

1. The solenoid

The solenoid constructed for this work was designed to fulfill the following requirements.

(a) The variation of the field strength over a light source 6 cm long, at the center of the coil must not exceed 0.1 percent.

(b) A method of cooling must be provided to permit continuous operation at full power.

(c) Subject to the demands of (a) and (b), the maximum possible field strength should be obtained from the available power supply.

Fig. 1 shows a cross section of the solenoid as it was built in the laboratory shop. The winding is continuous and consists of 2449 turns of No. 4 (5.2 mm) square, cotton-covered, copper wire in 18 layers. The coil proper is 80 cm long and has an outer diameter of 39.7 cm, and an inner diameter of 7.6 cm. The coil was wound on a heavy brass inner tube, between cast brass spiders. It was insulated from the tube by a layer of 3/16'' micarta strips laid longitudinally, and from the spiders by micarta strips attached to them. Between each two layers of the coil was placed a layer of black fiber spacers, 6.5 mm × 3.2 mm × 80 cm, parallel to the axis of the tube and spaced so as to

[9] See for example, Pauling and Goudsmit, "Structure of Line Spectra," p. 142.

leave passages through the length of the coil. A similar layer of spacers insulated the wire from the brass tube into which the completed coil was forced. The ends of the shell were closed by cast brass plates which were screwed to the inner tube and to a flange on the outer tube. The entire assemblage weighed 1200 lbs., and was supported by a wooden platform. In order to obviate the accidental presence of iron in the spiders and the end plates they were cast from freshly alloyed copper and zinc.

The solenoid was cooled by pumping kerosene through the passages left between the fibre spacers. A centrifugal pump maintained an estimated flow of about 200 liters per minute through a circuit consisting of the solenoid and eight automobile radiators. The radiators were assembled in a unit and immersed in a tank of running water. To protect the oil against contamination in case of leaks, the radiators were placed on the discharge side of the pump, so that the pressure of the kerosene was always higher than that of the water

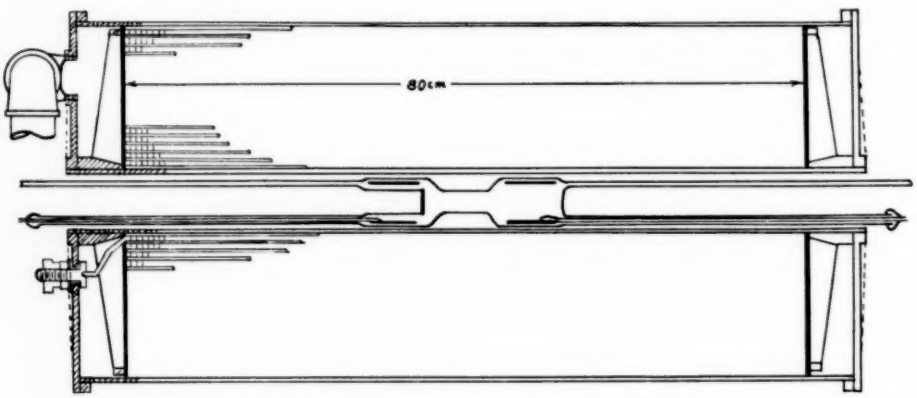


Fig. 1. Construction of solenoid.

outside. To test the insulation offered by the kerosene we frequently measured the resistance between the winding and the shell, and found it always between 10^5 and 10^6 ohms. The leakage resistance across the coil was necessarily of the same magnitude. As the resistance of the coil was 1.3 ohms between the terminals, the difference between the measured current and the effective current was negligible.

Two compound generators connected in series supplied the solenoid with a full-load current of over 200 amperes at 270 volts. The field circuit of one of the generators was controlled by means of a rheostat near the solenoid. A field of 7300 gauss, which required 54 kilowatts, could be maintained continuously without exceeding a temperature of 50°C in the circulating kerosene. During an exposure the exciting current was measured by means of a 0.001 ohm Leeds and Northrup shunt and a deflection potentiometer, and was controlled by adjusting the generator field current.

It was at first expected that it would be possible to determine the field-current ratio of this solenoid simply by careful measurement of the winding. For this purpose, each layer was carefully measured as it was wound. The

result obtained from this measurement was $K = 36.68$ gauss per ampere. This result is interesting, however, only in the fact that it is 0.5 percent lower than the value obtained by measuring the field in the completed coil. The difference is in the right direction to be explained as the result of the compression of the inner layers by the tension of those wound over them.

2. The source of light

The light was produced by evaporating zinc and cadmium in the positive column of a direct current discharge through helium. The tube, which is shown in Fig. 1, was constructed entirely of quartz. The illumination could be confined to a constriction, 6 cm long and elliptical in cross section (1 cm $\times$3 cm), which was placed at the center of the solenoid. The electrodes were short sections of copper tubing 2 cm in diameter. A re-entrant window extended through the anode to within 2 cm of the end of the constriction, and the proportions of the tube were such that neither of the electrodes was visible from the lens used to focus the light on the slit of the spectrograph. Then by adjusting the pressure of helium so that the cathode glow was confined to within a few mm of the cathode, no light entered the spectrograph except that emitted from the region of maximum magnetic field.

The helium was continuously circulated and purified by means of charcoal and liquid air. The heat of the discharge was sufficient to vaporize zinc and cadmium shavings placed in the constriction or in the cathode. The whole tube was wrapped with copper foil and asbestos to maintain a relatively uniform temperature throughout. At currents above 600 m.a. the spectra of zinc and cadmium were produced with such intensity as to completely suppress the helium lines. All exposures were made, however, at a lower current, since the zinc and cadmium lines were somewhat sharper when their intensities were roughly one-third of that of He λ5876. The current for the discharge was furnished by four 500 volt generators connected in series, and was regulated by a series resistance.

3. Optical apparatus

The Fabry-Perot interferometer was the one previously described in connection with the study of the hydrogen fine structure,[10] and was placed between the collimator and the prism. The plate surfaces were sputtered with gold which was sufficiently dense to show about twenty visible reflections of the filament of a ten watt lamp. If the resolving power is roughly estimated as half the number of visible reflections multiplied by the order of interference, it was about 10^6, since most of the exposures were made with an order of interference something over 10^5. This corresponds to a plate separation of 3.5 cm. The spectroscopic apparatus was mounted on a heavy concrete slab supported by twelve tennis balls. This arrangement proved stable, and effectively protected the interferometer from the unavoidable vibration of the cooling pump. The supports of the collimator, interferometer, and camera

[10] W. V. Houston, Astrophys. **64,** 81 (1926).

were independently fastened to the concrete base, so that each part could be separately alligned. To protect the interferometer and prism from temperature variations a tight wooden box was used, from which the slit and the camera projected through felt gaskets. The box rested on a felt pad covering the concrete base and could be removed without disturbing the interferometer. The temperature was observed with a Beckman thermometer and was regulated by electrical heating to within 0.05°C. The spectroscopic equipment was placed 2 meters from the solenoid to prevent disturbance either of the interferometer or of the solenoid field.

The aperture of the interferometer was reduced to about 1 cm by means of a diaphragm, and the necessary exposure times ranged from 30 to 90 seconds with Ilford Extra Rapid Panchromatic plates.

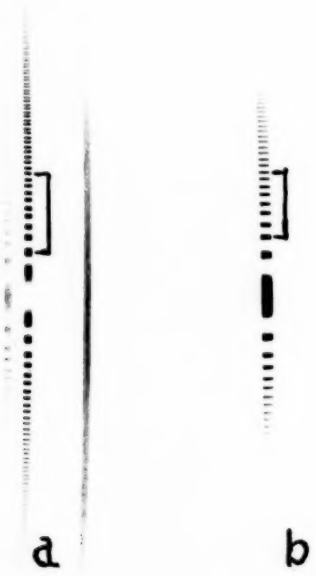


Fig. 2. Interferometer patterns. The brackets indicate pairs of fringes between which the Zeeman separation was measured.

Care was always taken to see that the fringes due to one component of the doublet lay half way between those due to the other. This prevents any displacement of the maxima due to slight overlapping. At the order of interference used, the normal Zeeman separation, with the maximum field of 7300 gauss, caused the components to overlap with a separation of 4.5 orders. In Fig. 2a the Zn and Cd red lines are shown at this separation. Two components, between which the separation would be measured, are indicated with a bracket. Because of the uneven spacing, this exposure was not measured. Fig. 2b shows the pattern of the Cd red line with the components separated approximately 3.5 orders at a field of 6860 gauss.

Magnetic Measurements

The intensity of the magnetic field during an exposure was determined by the relation

$$H = KI = KP/R \tag{4}$$

where K is the constant of the solenoid in gauss per ampere and I is the current in amperes. This current was determined from the reading P of a potentiometer connected across a shunt of resistance R. The measurements were made in international amperes, and the small correction to absolute units was made in the final result. The value of K was determined by comparing the field produced by a measured current in the large solenoid, with the field in a single layer standard solenoid. The magnetic constant of the standard solenoid could be calculated from its dimensions.

The comparisons between the standard solenoids and the large solenoid were made by two methods. The first is a null method, which affords a direct comparison between the two constants. It is limited, however, to measurements in which I is less than one ampere. On account of the possible influence of ferro-magnetic surroundings, heating, and slight distortion of the coil due to magnetic forces, it is undesirable to place too much dependence on the constancy of K over a large range of currents. Hence another method was also used which permitted determinations of K to be made for all currents.

1. The standard solenoids

Three different standard solenoids were used in the calibration. Two of them consisted of a layer of No. 12 bare copper wire wound on a bakelite tube which had been threaded with ten turns per inch. The bakelite was of linen stock in order to avoid the reputed ferromagnetism of paper stock bakelite. The other solenoid was wound with No. 20 enameled wire on a brass tube which had been threaded with 28 turns per inch, and upon which a layer of insulating varnish had been baked. All were of such a size that they could be placed within the inner tube of the large solenoid.

The number of turns per cm in the winding of each solenoid was determined by measurement with one of two scales. One was a Starrett steel meter, and the other was a glass scale taken from a cathetometer. Both scales were calibrated against a Gaertner type M1001 standard meter at Pomona College. We are indebted to the members of the physics department at Pomona, for their kind cooperation in helping us to make this comparison. At a temperature of 20.5°C the glass scale showed an excess length of 0.032 percent and the steel scale an excess of 0.008 percent. The error was uniform as nearly as could be determined, and has been applied to all measurements made with these scales. The measurements were made by placing the scale next to the solenoid, and reading the points at which a coincidence was observed between a scale division and the edge of a turn. The turn density in the various intervals was then evaluated and averaged over all the intervals. In no case was an appreciable deviation from uniformity noticed.

TABLE I. *Measurements of the standard solenoids.*

A. Dimensions, for the end correction.

Solenoid	Effective diameter in cm	Effective length in cm	End correction (cos α)
Bakelite No. 1	6.00	89.7	$0.99776 \pm 0.005\%$
Brass	5.86	90.0	$0.99788 \pm 0.002\%$
Bakelite No. 2	5.97	89.49	$0.99777 \pm 0.001\%$

B. Measurements of turn density, showing change with time.

Solenoid	Date	Scale	Number of readings	Number of intervals	Turns per cm	K_s	Mean deviation
Bakelite No. 1	10/10/30	Steel	20	10	3.93219	4.93028	0.02%
	5/ 4/31	Glass	32	70	3.93190	4.92990	0.01%
	7/13/31	Glass	12	36	3.92814	4.92519	0.01%
	8/29/31	Glass	38	116	3.92710	4.92389	0.01%
Brass	4/ 5/31	Steel	8	4	11.0236	13.8233	0.003%
	5/14/31	Glass	12	36	11.0235	13.8231	0.003%
	8/30/31	Glass	26	97	11.0230	13.8225	0.002%
Bakelite No. 2	7/ 8/31	Glass	66	174	3.93517	4.93406	0.03%
	8/26/31	Glass	12	36	3.93204	3.93014	0.02%

Table I gives the data from the measurements of the two solenoids together with the values of the constants computed from the relation

$$K_s = 0.4\pi n \cos \alpha \tag{5}$$

where n is the number of turns per cm and α is the angle subtended at the center by the radius at one end.

2. The null method calibration

The arrangement of apparatus for determining the ratios between the constant of the large solenoid and those of the standards is shown in Fig. 3.

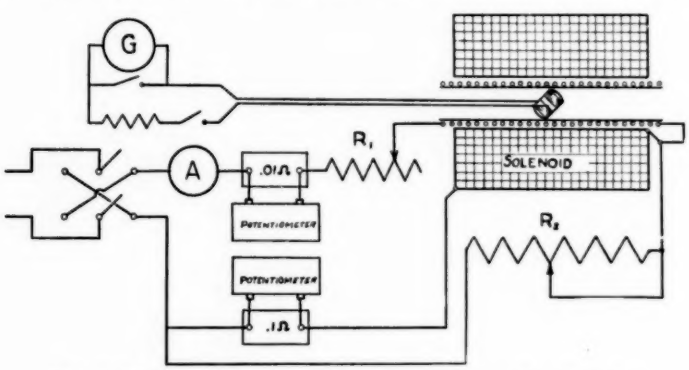


Fig. 3. Diagram of electrical connections for null calibration.

The standard solenoid was placed within the inner tube of the large solenoid and was connected so that the two fields were opposing. A large flip coil

wound with 10,000 ohms of No. 40 wire was placed at the common center of the solenoids and connected to a Leeds and Northrup wall-type ballistic galvanometer. The ratio of the currents in the two solenoids was then varied until a balance was indicated by a zero deflection in the galvanometer when the coil was turned over. The ratio of the constants is then given by

$$R = K/K_s = I_s/I \tag{6}$$

where the subscript $_s$ denotes the standard. The currents were measured with potentiometers and shunts.

In practice the currents were read at a series of values giving small galvanometer deflections. A plot of the deflections against the current ratios gave the balance point. The influence of the earth's field was removed by reversing both currents. The potentiometers were interchecked and showed a maximum departure of 0.02 percent. The same standard cell was used with both potentiometers. The shunts used were checked by the Southern California Edison Company's testing laboratory against resistances certified by the Bureau of Standards. In addition, the 0.1 ohm and the 0.001 ohm shunts were tested by the Bureau of Standards. The values of the resistances are given in Table II. The 0.1 ohm shunt was usually used at about 6 amperes so that the average of the values for 1.5 and 15 amp. has been used. The 0.001 ohm shunt was used at about 200 amp. At this current the heat developed is about half that at 300 amp. so the average of the two values was taken as being about correct.

TABLE II. *Calibration of shunts used.*

Nominal resistance (Ohms)	Edison Co. value	Bureau of Standards	Adopted value	Correction to be added to nominal value
		60 amp.		
0.001	0.00100050	0.0010004	0.0010000	0.000%
		300 amp.		
		0.0099994		
0.01	0.0099941		0.0099941	−0.059%
		1.5 amp.		
0.1	0.100024	0.10004	0.100035	+0.035%
		15 amp.		
		0.10003		
1.0	0.99955		0.99955	−0.045%
10.0	9.9969		9.9969	−0.031%

Table III gives the results of the null method calibration. There is no evidence of systematic change of the large solenoid between the summers of 1930 and 1931, so that all the observations were averaged together to get the final result. Each value of K in the last column was given a weight equal to the number of individual complete determinations from which it came. This number is given in column 3. The mean deviation is about one part in three thousand. During the summer of 1931 both bakelite standards changed considerably as can be seen from Table I. The value of K_s in Table III was obtained on the assumption that the change was linear with the time, at least between the times at which the solenoid was measured.

Since the current used in the standard solenoids caused them to become somewhat warmer than they were when measured, a small temperature cor-

TABLE III. *Calibration by the null method.*

Standard solenoid	Date	Number values	Mean P_s/P	Shunt corr.	I_s/I	K_s (Table I)	K
Bakelite No. 1	7/29/30	5	7.4766	+0.094%	7.4836	4.9302	36.896
	3/ 8/31	1	7.4739	−0.080%	7.4679	4.9300	36.817
	8/ 1/31	3	7.4899	+0.014%	7.4909	4.9247	36.890
Brass	8/ 9/31	3	2.6671	+0.014%	2.6675	13.823	36.873
	8/14/31	1	2.6701	−0.069%	2.6683	13.823	36.884
Bakelite No. 2	7/22/31	7	7.4781	−0.080%	7.4721	4.9328	36.858
	7/29/31	2	7.4769	+0.014%	7.4799	4.9325	36.887
	8/22/31	2	7.4830	−0.080%	7.4770	4.9305	36.865

Weighted Mean $K=36.872$ Temperature correction $=-0.011$ K Corrected $=36.863$
Mean deviation $=0.018$

rection must be applied to these values. The rise in temperature was about 15°, so that the correction to be applied is −0.03 percent. This is no more than the mean deviation of the various measurements.

3. Calibration with a mutual inductance.

In order to permit determinations of the solenoid constants to be made under conditions similar to those prevailing during an exposure, the arrangement shown in Fig. 4 was adopted. The current in the primary of the mutual

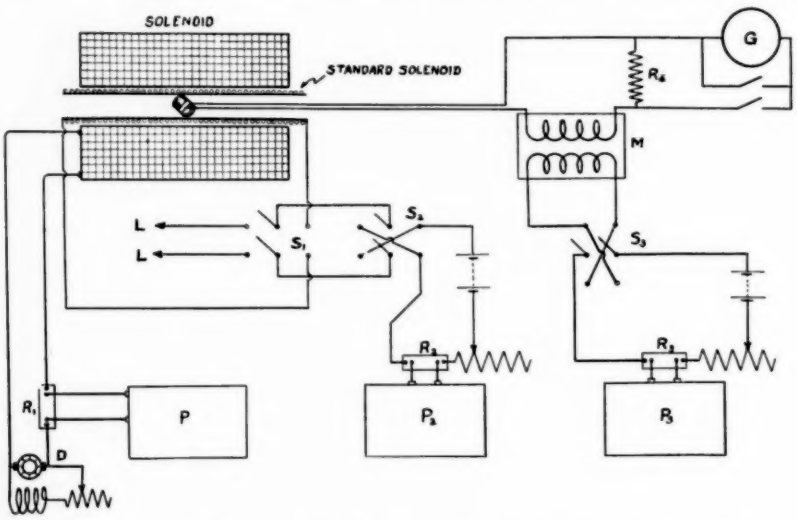


Fig. 4. Diagram of connections for calibration with mutual inductance.

inductance, M, was varied until its reversal gave a deflection of the galvanometer equal to that produced when the flip coil was operated in the field of the standard solenoid alone. The galvanometer shunt was such that full scale deflections were obtained. Since the total resistance in the galvanometer cir-

cuit was the same for both deflections, we had, in effect, a calibration of the mutual inductance and the flip coil in terms of the standard solenoid constant and the ratio of the currents. Let K_s = constant of the standard solenoid, gauss per ampere, I_s' = current in standard solenoid in amperes, $H_s = K_s I_s'$ in gauss, F = magnetic area of the flip coil, M = mutual inductance, I_m' = current in mutual inductance.

Then $2FH_s = 2MI_m'$ and $FK_s I_s' = MI_m'$, whence

$$M/F = K_s I_s'/I_m'. \tag{7}$$

After calibration, the mutual inductance could be used to determine the constant of the large solenoid at full current. The galvanometer shunt R_4 was increased until a suitable deflection was obtained when the flip coil was operated in the field of the large solenoid. The current in the primary of the mutual inductance was then raised until a reversal of this current with the switch S_3 gave the same deflection. The currents I and I_m in the solenoid and the mutual inductance respectively were then read from the potentiometers. Then

$$K = MI_m/FI \tag{8}$$

With Eq. (7) this gives

$$K = K_s I_s' I_m/I_m' I. \tag{9}$$

The accuracy and reliability of this method are dependent on several factors, which may be listed as follows:

(a) The galvanometer constant must not change during the calibration of the inductance, nor during the calibration of the solenoid from the inductance. The galvanometer used was a Leeds and Northrup Type P ballistic instrument. It was found necessary to found necessary to place it at least 70 feet from the solenoid, in order that the stray field would not change the constant. Although the galvanometer reading had to be reproducible, almost no dependence was placed upon the proportionality of the deflection to the charge which passed through the secondary. A series of deflections, covering a small range, was made using the flip coil. Intermingled with these was a second series of readings, in the same small range, obtained by reversing the current in the primary of the mutual inductance. Each deflection was then plotted against the current at which it was read. Graphical interpolation to the same deflection in both series gave the desired current ratio, I_m/I or I_m'/I_s'.

(b) The method also requires that the mutual inductance have the same value when being compared with the large solenoid as it has during the comparison with the standards. This requires constancy of inductance over a wide range of primary current values. Comparisons with the standards were intermingled with comparisons with the large solenoid so that any permanent change in inductance would show itself in the results. No such change was observed as can be seen from Table IV. Errors due to ferromagnetic surroundings were minimized by suspending the inductance midway between the floor and the ceiling of a large adjacent room. The construction of the inductance was such as to guard against possible changes with time or current.

A low resistance primary of No. 12 wire was wound on a micarta tube 8 inches in diameter and 15 inches long. The secondary consisted of 12 lbs. of No. 32 S.C.C. and enameled wire wound in three separate coils and rigidly supported inside the primary.

TABLE IV. *Calibration of the mutual inductance.*

	No. 1 Bakelite solenoid	Brass solenoid
Measured P_s/P_m	39827	14225
	39834	14234
	39842	14230
	39834	14223
	39851	14234
	39829	14227
	39840	
	39813	
	39846	
	39852	
	39854	
	39833	
Shunt	39838 ± 8	14229 ± 4
Correction -0.066%	26	$+0.014\%$ 2
	39812	14231
After temperature correction K_s	4.9284	13.819
K_sI_s'/I_m'	1962.09 ± 0.4	1966.58 ± 0.2
Weighted mean from the two solenoids		$K_sI_s'/I_m' = 1963.59 \pm 2.0$

(c) The required constancy of the flip coil was indicated by the coherence of the readings.

(d) The galvanometer shunt resistance needed to remain constant only during each half of the calibration. Separate coils of Chromel wire were prepared for each required value of this resistance.

Table IV gives the results of the calibration of the mutual inductance against the standard solenoids. The surprising thing is the large difference between the results obtained with the two different standards. This is far larger than the accidental errors in the measurements, and is far larger than one has a right to expect, from the results of the null method calibration. This unexplained discrepancy indicates that the uncertainty is at least 0.1 percent.

Table V gives the results of the calibration of the large solenoid against the mutual inductance. Readings were taken for three different ranges of current. One was in the neighborhood of 1 amp., the next between 100 and 125 amp., and the third was near 200 amp. The slight systematic trend shown by the results is less than the probable uncertainty of the readings and so has been neglected. The mean of these three determinations was averaged with the mean of the null method readings to get the final value of the solenoid constant.

In this calibration three potentiometers were used which were checked against each other. The maximum disagreement was 0.03 percent while the average was about 0.001 percent. All potentiometers were always used with

TABLE V. *Calibration of large solenoid against mutual inductance.*

Current range	200 amp.	100 amp.	1 amp.
Observed	53282	53278	53266
Potentiometer	53356	53322	53315
Ratios	53311	53311	53289
P/P_m	53311	53311	53297
	53305		53285
	53289		
	53262		
	53282		
	53300		
	53283		
	53299		
Mean	53298 ± 17	53305 ± 14	53290 ± 12
Shunt correction +0.035%	19 +0.015%	08 +0.014%	7
I/I_m	53316 ± 17	53313 ± 14	53297 ± 12
$K=\frac{K_s I_s' I_m}{I_m' I}$	= 36.829 ± 0.041	36.831 ± 0.041	36.842 ± 0.041
Mean value of K = 36.834 ± 0.041			

the same standard cell, so that no corrections were necessary and small changes of the cell with time were of no account.

Although the calibration with the mutual inductance shows some four times the uncertainty of the null method calibration, the two values have been averaged with equal weight, since the calibration under actual operating conditions is rather to be preferred to the null method where only low currents could be used. This gives for the adopted value

$$K = 36.846 \text{ } \textit{gauss per ampere}.$$

This value has been reduced by 0.005 percent from the mean of Tables IV and V to change the units from international to absolute gauss per ampere.

SPECTROSCOPIC MEASUREMENTS

In terms of the solenoid constant and the current the expression (3) for e/m is

$$e/m = (4\pi c/\bar{a}K)(\Delta\nu/I). \tag{3'}$$

The difference between the wave numbers of the two Zeeman components is found from the difference in order of interference at the center of the fringe pattern.

The relation for the fractional order of interference at the center of the pattern, in terms of the diameters of the fringes, is

$$p = \frac{D_i^2}{D_i^2 - D_{i-1}^2} - i \tag{10}$$

where D_i is the linear diameter of the i-th ring from the center of the pattern. Two approximations are involved in Eq. (10). Tan θ has been substituted for θ, and $1-\theta^2/2$ has been substituted for cos θ. The errors thus introduced are quite negligible since θ, the angular diameter of the largest fringe meas-

ured, was less than 0.04 radians. The error is further reduced by the fact that only the differences between values of p are used.

The diameters, D_i, of roughly twenty fringes of each component were measured on a comparator. From the table of values of D_i^2 a mean value of $(D_i^2-D_{i-1}^2)$ was obtained. Dividing this into each D_i^2 gave, by Eq. (10), a table of $(i+p)$ for each component. The mean of the fractional parts was then the desired order of interference at the center. From the difference in order, $(i+p)-(i'+p')$, between the fringes corresponding to $+\Delta\nu$ and $-\Delta\nu$ in the Zeeman pattern, the separation in cm^{-1}, or Balmers, is given by

$$2\Delta\nu = (i + p - i' - p')/2dn \qquad (11)$$

where n is the index of refraction of air. This use of the index of refraction of air is necessary since the value of the velocity of light in vacuum is used. The method of Lord Rayleigh was used in evaluating d which could be determined with an accuracy of one or two parts in a million.[11] The order of interference was changed after every three or four exposures.

During the exposures the solenoid current was measured by the same shunt and potentiometer that were used in the 200 ampere calibration. By constant regulation of the generator field the current was maintained with average fluctuations of less than 0.1 percent. A week before the exposures were started, the standard cell was checked with a new cell having a Bureau of Standards certificate. Several checks indicated that the cell was constant and known to at least one part in ten thousand.

TABLE VI. *Spectroscopic measurements.*

Cadmium 6439				Zinc 6362			
Plate	$2\Delta\nu/I$	Order diff.	Wt.	Plate	$2\Delta\nu/I$	Order diff.	Wt.
1	0.0034386	1.4587	1	9	0.0034379	4.5130	3
2	344	4.5109	2	11	78	4.5127	1
3	412	4.5195	2	12	66	4.4882	2
4	382	4.5157	2	14	74	4.5132	2
5	416	3.5122	2	16	70	4.5127	1
6	355	4.5082	2				
7	338	4.4819	1				
8	402	3.5080	1				
10	329	4.5088	1				
12	396	4.4921	1				
13	399	4.5190	2				
14	336	4.5080	2				
15	357	4.5108	2				
17	341	2.4877	1				
Mean	0.0034372 ± 28				0.0034374 ± 5		
$2\pi c/aK'$	51149				51134		
e/m	1.7581 ± 0.0025				1.7577 ± 0.0025		

Thirty-one separate reductions of nineteen interferometer patterns gave the values of $2\Delta\nu/I$ listed in Table VI. More than half of the patterns were independently measured and reduced by both of us. In these cases the results showed an average discrepancy of 0.03 percent and a maximum discrepancy

[11] Rayleigh, Phil. Mag. **9**, 685 (1906).

of 0.13 percent. The mean of each pattern is given a weight equal to the number of times it was measured. In determining e/m the value of c was taken as 2.99796×10^{10} cm per sec. K' is the value of K adopted above, but reduced by 0.016 percent to take account of the fact that the field in the solenoid was not entirely uniform and had to be averaged over the light source.

As a result of these measurements we may say with fair certainty that the spectroscopic value is

$$e/m = 1.7579 \pm 0.0025 \times 10^7 \text{ e.m.u. per gram.}$$

This is rather lower than the previous spectroscopic values, although the accuracy is not great enough to make certain that it is really different. However it is certainly far lower than the high deflection values.

DIRAC'S EQUATION AND THE SPIN-SPIN INTERACTIONS OF TWO ELECTRONS

By G. Breit
Department of Physics, New York University

(Received January 6, 1932)

Abstract

An improvement is made on a previous attempt to treat two particles by means of Dirac's equation. The approximate equation (1) below is considered in successive steps. The first step, following Oppenheimer, includes the electrostatic energy exactly, rather than to the first in power in e^2. This makes it possible to use it as a good starting point in the calculation of spectral terms. The second step brings in the energy due to the interaction of the electric currents. It is given by (9) below. Maxwell's equations and the conservation of energy (see (8.5)) demand the validity of the diagonal matrix elements of this expression as a first order perturbation energy, independently of theories of light quanta. The interactions of the particles with themselves give *additive constants* in the energy within the limits of the approximation used. Within the same limits the results are in agreement with experiment.

(I). Statement of Problem

IN A previous paper[1] an attempt has been made to treat the spin-spin interactions of two electrons by means of Dirac's relativistic equation.[2] The conclusion reached at the time was that the equation

$$\{p_0 + \mathbf{a}^{\mathrm{I}}\mathbf{p}^{\mathrm{I}} + \mathbf{a}^{\mathrm{II}}\mathbf{p}^{\mathrm{II}} + (\alpha_4^{\mathrm{I}} + \alpha_4^{\mathrm{II}})mc + (e^2/2c)[\mathbf{a}^{\mathrm{I}}\mathbf{a}^{\mathrm{II}}r^{-1} + (\mathbf{a}^{\mathrm{I}}\mathbf{r})(\mathbf{a}^{\mathrm{II}}\mathbf{r})r^{-3}]\}\psi = 0 \quad (1)$$

is not in agreement with experiment.[3]
Here

$$p_0 = -\frac{h}{2\pi i}\frac{\partial}{c\partial t} + (e/c)(A_0^{\mathrm{I}} + A_0^{\mathrm{II}}) - (e^2/cr)$$

$$\mathbf{p}^{\mathrm{I}} = (p_1^{\mathrm{I}}, p_2^{\mathrm{I}}, p_3^{\mathrm{I}}), \qquad \mathbf{p}^{\mathrm{II}} = (p_1^{\mathrm{II}}, p_2^{\mathrm{II}}, p_3^{\mathrm{II}})$$

$$p_k^{\mathrm{I}} = \frac{h}{2\pi i}\frac{\partial}{\partial x_k^{\mathrm{I}}} + (e/c)A_K^{\mathrm{I}}, \quad p_k^{\mathrm{II}} = \frac{h}{2\pi i}\frac{\partial}{\partial x_k^{\mathrm{II}}} + (e/c)A_K^{\mathrm{II}}, \quad (k = 1, 2, 3)$$

(A_0, A_1, A_2, A_3) = electromagnetic four vector
$r = +[(x_1^{\mathrm{I}} - x_1^{\mathrm{II}})^2 + (x_2^{\mathrm{I}} - x_2^{\mathrm{II}})^2 + (x_3^{\mathrm{I}} - x_3^{\mathrm{II}})^2]^{1/2}$ = distance between electrons I and II $\mathbf{a}^{\mathrm{I}} = (\alpha_1^{\mathrm{I}}, \alpha_2^{\mathrm{I}}, \alpha_3^{\mathrm{I}})$, $\mathbf{a}^{\mathrm{II}} = (\alpha_1^{\mathrm{II}}, \alpha_2^{\mathrm{II}}, \alpha_3^{\mathrm{II}})$
Matrices $(\alpha_1^{\mathrm{I}}, \alpha_2^{\mathrm{I}}, \alpha_3^{\mathrm{I}}, \alpha_4^{\mathrm{I}})$, $(\alpha_1^{\mathrm{II}}, \alpha_2^{\mathrm{II}}, \alpha_3^{\mathrm{II}}, \alpha_4^{\mathrm{II}})$, identical in form with Dirac's

[1] Breit, Phys. Rev. **34**, 553 (1929).
[2] An attempt along the same lines has been also made by Gaunt (Phil. Trans. Roy. Soc. **A228**, 151 (1929). A discussion of this will be found in references I, 3.
[3] Breit, Phys. Rev. **36**, 385 (1930).

s, are operators on two independent indices belonging respectively to I and II, each index taking four values 1, 2, 3, 4.

With a slight modification it has been shown however to fit the observed facts as well as could be expected. The modification required consists in the removal of terms in e^4 in the result of eliminating the "small" components of the wave-function. These terms do not involve h so that their physical significance is doubtful even for this reason alone.

Eq. (1) is not exact. It is wrong when the distance between the particles or their velocities are too great. It also was not intended to give account of radiation from the atom. It will now be shown that a modification of the first interpretation of (1) can be given in such a way that the difficulty with the e^4 terms disappears. From the present point of view the e^4 terms derived in reference 1 have no direct physical significance. Their appearance in the previous work was due to trying to attach more exactness to the last term in (1) than it possesses.

In reference 1 the derivation of the equation has been given from the point of view of the Heisenberg-Pauli theory of electrodynamics. It was supposed in this calculation that the interaction energy between the particles is small. The first order effect in the interaction energy was calculated and it was found that the values of the interaction integrals to be used in the calcalculation of the mutual energy are given by the formula (63) (using umbral notation)

$$A_{st,ts} = \int [(u_\rho^{*s} u_\rho^t)\{r_{PP'}^{-1}\}(u_\sigma^{*t} u_\sigma^s)' - (u_\rho^{*s} \alpha_{\rho\sigma} u_\sigma^t)\{1/r_{PP'}\}(u_\mu^{*t} \alpha_{\mu\nu} u_\nu^s)' dV dV' \qquad (2)$$

$$\{1/r_{PP'}\} = (1/r_{PP'}) \cos(2\pi r_{PP'}/\lambda_{st}).$$

Here s, t are unperturbed electron levels. The wave-length of the transition $s \rightarrow t$ is denoted by λ_{st}. Recently, C. Møller[4] and L. Rosenfeld[5] have given simpler derivations of the same result also to the first order in e^2 without using the formalism of the quantum electrodynamics.[6] There is hardly any doubt that the mutual energy of two electrons is given correctly to the first order in e^2 by the interaction integrals $A_{st,ts}$, as long as one is justified in speaking of a constant energy.

This derivation is not sufficient to make practical applications to the fine structure of He possible. The Shroedinger non-relativistic wave equation gives, according to Kellner and Hylleraas as well as others, term value energies for the two electron problem which are in good agreement with experiment. In these calculations the electrostatic energy e^2/r is used much more accurately than to the first order. It is desirable to have such a theory that Schroedinger's non-relativistic wave equation with the e^2/r term for the mu-

[4] C. Møller, Zeits. f. Physik **70**, 786 (1931).

[5] L. Rosenfeld, Zeits. f. Physik **73**, 253 (1931).

[6] The fact that the easiest interpretation of $A_{st,ts}$ is in terms of the retarded potentials has been stated in the writer's paper. (p. 572).

tual energy could be taken as a starting point for the calculations. In order to have such a theory it is necessary to explain the validity of the e^2/r term in an exact sense. We must strive therefore for a theory in which the Schroedinger non-relativistic value of the energy is modified by the presence of small terms.

II. Electrostatic Energy

The proper point of view for such a treatment is given by Oppenheimer[7] making use of the second form of the Heisenberg-Pauli theory. It is shown in Oppenheimer's paper that the variables describing the longitudinal components of the electromagnetic field can be eliminated from the wave equations. (Formula (8) of J. R. Oppenheimer.[7]) The process of elimination brings in infinite terms into the energy expression. The infinite terms represent the infinite electrostatic energy of point charges. It is stated in Oppenheimer's paper that these terms may be considered as constants and do not interfere with the application of the theory. This point is of special interest to us and will therefore be discussed somewhat more in detail.

The mathematical apparatus of Heisenberg-Pauli II can be applied to classical electrodynamics. Its direct application also leads to infinities as would be expected of any theory with point charges. Nevertheless, in classical theory, there is a way of escaping the difficulty. In all of the expressions for the interaction energy we can replace $\boldsymbol{A}(\boldsymbol{r})\dot{\boldsymbol{r}}$ by $\int \boldsymbol{A}(\boldsymbol{r}')D(\boldsymbol{r}-\boldsymbol{r}')\dot{\boldsymbol{r}}dV'$ where $\boldsymbol{A}$ is the vector potential, $\dot{\boldsymbol{r}}$ the velocity, $\boldsymbol{r}$ is the position of the particle and $\boldsymbol{r}'$ is a variable point in space. The function D is a concentrated function satisfying

$$\int D(\boldsymbol{r}-\boldsymbol{r}')dV' = 1$$

resembling a δ function before the latter passes to the limit of infinite concentration. Such a scheme can be carried through only non-relativistically in a classical particle theory. Nevertheless the result is of interest because of its simplicity. The longitudinal components of the field can be eliminated by using the equation div $\mathcal{E}=4\pi\rho$. The elimination introduces additional terms into the Hamiltonian. Their sum is simply the energy of the electrostatic field (defined by div $\mathcal{E}=4\pi\rho$, curl $\mathcal{E}=0$) of the charge distributions described by $D(\boldsymbol{r}-\boldsymbol{r}')$ for each particle. It may be decomposed into a sum of the electrostatic energy of the field which would exist if the particles were infinitely far from each other and the mutual energy of the particles ($\Sigma_{i,j}e_ie_j/r_{ij}$). The self energy is constant and may be removed. Similarly in quantum theory we can replace all integrals

$$\int A(\boldsymbol{r})(\psi_\mu^*\alpha_{\mu\nu}\psi_\mu)dV \text{ by } \int\int A(\boldsymbol{r}')D(\boldsymbol{r}-\boldsymbol{r}')(\psi_\mu^*\alpha_{\mu\nu}\psi_\nu)_{\boldsymbol{r}}dVdV'.$$

We then obtain large additive constants *identical with those of the classical theory*. The additive constants can be removed from the wave equation with-

[7] J. R. Oppenheimer, Phys. Rev. **35**, 461 (1930).

out affecting the physically interesting conclusions about numbers of light quanta, distribution of particles, etc. The fact that a hydrogenic atom may be treated by means of Dirac's equation shows that this removal of the additive constants is a safe procedure. From our present point of view there is no essential difference between the mutual electrostatic energy of a proton and an electron and the mutual electrostatic energy of a collection of several protons and several electrons.

There is an apparent objection. As long as the D function is not the δ function the equations are not relativistically invariant. The results however should approach relativistically invariant results as D approaches δ.

On elimination we end up therefore with Oppenheimer's (11) slightly modified by the removal of the large additive constants and by the replacement of every e_ie_j/r_{ij} by the mutual electrostatic energy of two charge distributions $e_iD_i(\boldsymbol{r}_i-\boldsymbol{r}')$ $e_jD_j(\boldsymbol{r}_j-\boldsymbol{r}')$, $r_{ij}=|\boldsymbol{r}_i-\boldsymbol{r}_j|$. This is then identical in content with Dirac's theory of light quanta.[8]

III. Interactoin with Transverse Waves

In Dirac's theory of light quanta part of the electromagnetic field is described by means of a vector potential $\boldsymbol{A}$, satisfying div $\boldsymbol{A}=0$. In order to describe the whole field we need in addition an electrostatic field. We call it $\mathcal{E}'$ and the total electric field we call $\mathcal{E}$. The defining equations are then

$$\mathcal{E} = -\frac{\partial \boldsymbol{A}}{c\partial t} + \mathcal{E}';\ \mathfrak{H} = \text{rot } \boldsymbol{A};\ \text{div } \mathcal{E}' = 4\pi\rho;\ \text{rot } \mathcal{E}' = 0 \tag{3}$$

$$\text{div } \boldsymbol{A} = 0.$$

Here ρ is the charge density supposed to satisfy the continuity equation

$$\frac{\partial \rho}{\partial t} + \text{div } j = 0 \tag{4}$$

where j is the current density. It follows from Maxwell's equations that

$$\left(\Delta - \frac{\partial^2}{c^2\partial t^2}\right)\boldsymbol{A} = -\frac{4\pi j}{c} - \frac{\partial \mathcal{E}}{c\partial t}. \tag{5}$$

By (3) and (4)

$$\left(\frac{\partial \mathcal{E}'}{\partial t}\right)_P = \nabla_P \int \frac{(\text{div}_{P'}\, j_{P'})}{r_{PP'}}\, dV_{P'} \tag{6}$$

where $r_{PP'}$ is the distance between points P and P'. Hence substituting (6) into (5) using the retarded expression for A we have

$$\boldsymbol{A}_{P''} = \frac{1}{c}\int \frac{\{j_{P'}\}}{r_{P'P''}}\, dV_{P'} + \frac{1}{4\pi c}\int \frac{1}{r_{P''P}}\left\{\nabla_P \int j_{P'}\nabla_P\left(\frac{1}{r_{PP'}}\right)dV_{P'}\right\}dV_P \tag{7}$$

[8] The obvious changes of using progressive waves in the Heisenberg-Pauli theory and of using Dirac's relativistic equation in Dirac's presentation of the theory of light quanta must be made in order to obtain also identity in form.

where $\{\}$ denotes retarded values. Neglecting retardation in (7) we obtain after some easy transformations

$$\int j_{P''}\boldsymbol{A}_{P''}dV_{P''} = \frac{1}{2c}\int\left(\frac{j_{P'}j_{P''}}{r} + \frac{(j_{P'}\boldsymbol{r})(j_{P''}\boldsymbol{r})}{r^3}\right)dV_{P'}dV_{P''} \qquad (8)$$

where r is the distance between points P', P''. The above deduction of (8) applies in the classical as well as in the quantum theory as long as the current operators commute with the coordinates. The operator which represents the charge density at a point $\boldsymbol{r}'$ is $-\Sigma e_i D_i(\boldsymbol{r}_i-\boldsymbol{r}')$ while the current density operator is given by $\Sigma e_i c\,\boldsymbol{\alpha}^i D_i(\boldsymbol{r}_i-\boldsymbol{r}') = -\Sigma e_i \dot{\boldsymbol{r}}_i D_i(\boldsymbol{r}_i-\boldsymbol{r}')$.

We now consider the wave equation which results on eliminating the electrostatic field:

$$\left(p_0 + \sum_i \boldsymbol{\alpha}^i\left(\boldsymbol{p}^i + \frac{e_i}{c}\int \boldsymbol{A}(\boldsymbol{r}')D_i(\boldsymbol{r}_i-\boldsymbol{r}')dV'\right)\right.$$
$$\left. + \sum_i \alpha_4{}^i m_i c - \frac{1}{8\pi c}\int\left[\left(\frac{\partial \boldsymbol{A}}{c\partial t}\right)^2 + \mathfrak{H}^2\right]dV\right)\phi = 0 \qquad (58).$$

$$\left(p_0 = -\frac{h\partial}{2\pi i c\partial t},\quad p_k{}^i = \frac{h\partial}{2\pi i \partial x_k{}^i}\right) \qquad (k = 1, 2, 3).$$

The wave function contains the coordinates of the particles as well as those describing the radiation (e.g., numbers of light quanta). We shall not use explicitly the way in which ϕ depends on the radiation variables. The term due to $(\partial A/c\partial t)^2$ will be neglected since it has to do with the acceleration. It cannot be taken into account consistently without considering the emission of radiation. We may now use expressions (8), (7) in the above wave equation. The energy operator is then expressed in terms of the particle variables without the aid of the radiation variables. The resultant contribution to the energy operator is

$$E = -\sum_i e_i\boldsymbol{\alpha}^i\int \boldsymbol{A}(\boldsymbol{r}')D_i(\boldsymbol{r}_i-\boldsymbol{r}')dV' + (1/8\pi)\int \mathfrak{H}^2 dV.$$

We have

$$\int \mathfrak{H}^2 dV = \int \mathfrak{H}\ \mathrm{rot}\ \boldsymbol{A}dV = \int(\mathrm{rot}\ \mathfrak{H})\boldsymbol{A}dV = \int\left(\frac{4\pi \boldsymbol{j}}{c} - \frac{\partial^2 \boldsymbol{A}}{c^2\partial t^2} + \frac{\partial \mathcal{E}'}{c\partial t}\right)\boldsymbol{A}dV$$

We neglect $\partial^2\boldsymbol{A}/c^2\partial t^2$ and $\partial\mathcal{E}'/\partial t$ being a gradient contributes nothing by Green's theorem, div $\boldsymbol{A}$ being zero. Thus

$$\Delta E = -\frac{1}{2c^2}\int j\boldsymbol{A}dV = -\frac{1}{4c^2}\int\left(\frac{(j_{P'}j_{P''})}{r} + \frac{(j_{P'}\boldsymbol{r})(j_{P''}\boldsymbol{r})}{r^3}\right)dV_{P'}dV_{P''}. \qquad (9)$$

This is equivalent to Oppenheimer's (28).[7] In the derivation of (9) we have made a number of approximations. These have been suggested to us by the belief that acceleration and radiation have little to do with the energy of sta-

tionary states. It is not possible to justify this belief by present mathematical theories. There is thus an element of speculation in the formula (9). We believe that it is correct to start with the wave equation omitting A, to use the energy eigenwerte of this equation and to add to them the diagonal matrix elements of (9). This should give the first order perturbation in the energy due to the interaction of the particles with the transverse electromagnetic waves. Again it is not possible to give a mathematical justification because there exists no proper quantum theory of electrodynamics. However, simple considerations indicate that the diagonal matrix elements of (9) are good approximations to the desired correction for the interaction with the field.

The point is that we may multiply (8.5) by ϕ^* on the left and integrate over the configuration space including that of light quanta. We then have an equation connecting the "expectations" (Erwartungwerte) of the terms in the operator of (8.5). This equation simply means that the "expectation" of the energy is the sum of the kinetic energy of the material particles and of the energy of the electromagnetic field. There is every reason to believe that this is correct even though present theories of electrodynamics based on (8.5) and special assumptions about the nature of light quanta lead to some physically impossible results. The coupling of the motions of particles due to their interactions with transverse waves is small compared to that due to electrostatic forces. We may therefore compute the effect on the "expectation" of the energy by using the unperturbed proper functions of (8.5) with A omitted. Thus the correctness of the diagonal elements of ΔE in energy calculations follows from the principle of conservation of energy and the form of (8.5) independently of the assumptions made about the nature of light quanta.

(IV) Constancy of Self-Energy Terms

Replacing the current operators by $\Sigma e_i c \mathfrak{a}^i D_i(\mathbf{r}_i - \mathbf{r}')$ we have for the addition to the energy characterised by quantum numbers $n_1, n_2, n_3 \ldots$ in the wave equation

$$\left(p_0 + \sum_i \mathfrak{a}^i \mathfrak{p}^i + \sum_i \alpha_4{}^i m_i c\right)\psi = 0 \tag{10}$$

the diagonal matrix element of the operator (9):

$$(\Delta E)_{n,n} = -\left(\tfrac{1}{4}\right)\sum_{i,j} e_i e_j \int\int\left[\left(\frac{\mathfrak{a}^i \mathfrak{a}^j}{r_{P'P''}} + \frac{(\mathfrak{a}^i \mathbf{r}_{P'P''})(\mathfrak{a}^j \mathbf{r}_{P'P''})}{r_{P'P''}}\right) D_i(\mathbf{r}_i - \mathbf{r}') \cdot D_j(\mathbf{r}_j - \mathbf{r}'')\right]_{n,n} dV_{P'} dV_{P''}. \tag{11}$$

According to the above reasoning this formula is independent of the detailed assumptions about the nature of light quanta which are made in Dirac's theory. The essential requirements for its validity are the correctness of Maxwell's equations, the legitimacy of neglecting retardation in (7) and of $(\partial^2 \mathbf{A}/c^2 \partial t^2)$, $1/c^2(\partial \mathbf{A}/\partial t)^2$ in comparison with $4\pi j/c$, $\mathfrak{H}^2$ respectively. It may be also derived by means of Dirac's theory of light quanta or its equivalent

Heisenberg-Pauli theory as has been done by Oppenheimer.[7] Carrying through the calculation by the method of successive approximations of wave functions used in Heisenberg-Pauli I for the evaluation of their E^2 there is no difficulty in deriving $(\Delta E)n$, n in agreement with (11). The treatment of resonance denominators is then exactly similar to that in HPI for e^2/r. This adds to our confidence in (9) to the first order in e^2.

The summation over i, j can be broken up into two parts one part containing terms with $i \neq j$ and another with $i=j$.

$$-\tfrac{1}{4}\sum_{i,j} = -\tfrac{1}{2}\sum_{i>j} - \tfrac{1}{4}\sum_{i=j}.$$

The first part is in agreement with (1) provided the last term in (1) is used for the calculation of the energy by the usual method of the perturbation theory. The second part contributes a constant. In fact

$$\frac{\mathbf{a}^i\mathbf{a}^i}{r_{P'P''}} + \frac{(\mathbf{a}^i\mathbf{r}_{P'P''})(\mathbf{a}^i\mathbf{r}_{P'P''})}{r_{P'P''}} = \frac{4}{r_{P'P''}}.$$

The contribution to (11) due to these terms is

$$-\sum_i e_i^2\left(\int\int \frac{D_i(\mathbf{r}_i - \mathbf{r}_{P'})D_i(\mathbf{r}_i - \mathbf{r}_{P''})}{r_{P'P''}}\, dV_{P'}dV_{P''}\right)_{n,n}.$$

The double integral is a constant and therefore the diagonal matrix element is simply this constant. We have thus a contribution

$$-\sum_i 2W_0{}^i\left(W_0{}^i = (e_i^2/2)\int\int \frac{D_i(\mathbf{r}_i - \mathbf{r}_{P'})D(\mathbf{r}_i - \mathbf{r}_{P''})}{r_{P'P''}}\, dV_{p'}dV_{p''}\right)$$

to the energy where $W_0{}^i$ is the energy of the electrostatic field due to the particle i. It is not very satisfactory to have this energy thrust upon us. However it must be remembered that only energy differences can be observed. For these the additive constant energy is not of interest. As long as $\Sigma_{i>j}$ is small the perturbation in the energy differences between various spectral terms due to (11) is also small.

We see therefore that Eq. (1) may be interpreted consistently in the sense that the interaction of the particles with the transverse waves brings about a perturbation in the energy of the amount

$$(\Delta E')_{n,n} = -\tfrac{1}{2}\sum_{i>j} e_ie_j\left(\frac{\mathbf{a}^i\mathbf{a}^j}{r_{ij}} + \frac{(\mathbf{a}^i\mathbf{r}_{ij})(\mathbf{a}^j\mathbf{r}_{ij})}{r_{ij}{}^3}\right)_{n,n}. \tag{12}$$

The unperturbed energy values are the eigenwerte of (10).

This conclusion is somewhat more restricted than the supposition that (1) is correct. We should substitute for the last term in (1) its diagonal matrix element referred to the proper energy values of (10). Another way of expressing the same result is to say that (1) may be used for energy calculations if the last term is taken into account to the first order in the usual way by penturbation theory working with 16 Ψ components.

(V) Comparison with Experiment

It remains to compare the results with experiment. It is convenient to use Dirac's original form of the α matrices. As in reference (1) we let σ be the negative of Pauli's matrices. We introduce a wave function Ψ having four components identical with the four "large" components of ψ. The σ's operate on Ψ. Thus

$$(\sigma^{\mathrm{I}}\Psi)_{\alpha,\beta} = \sum_{\alpha'} \sigma_{\alpha\alpha'}\Psi_{\alpha'\beta} \qquad (\alpha, \beta = 3 \text{ or } 4).$$

We let

$$\begin{aligned}
(X^{\mathrm{I}}\Psi)_{3,\alpha} &= \psi_{1,\alpha}(X^{\mathrm{I}}\Psi)_{4,\alpha} = \psi_{2,\alpha} \\
(X^{\mathrm{II}}\Psi)_{\alpha,3} &= \psi_{\alpha,1};\ (X^{\mathrm{II}}\Psi)_{\alpha,4} = \psi_{\alpha,2} \\
(Y\Psi)_{3,3} &= \psi_{1,1};\ (Y\Psi)_{3,4} = \psi_{1,2};\ (Y\Psi)_{4,3} = \psi_{2,1};\ (Y\Psi)_{4,4} = \psi_{2,2}.
\end{aligned} \tag{13}$$

Then (1) may be rewritten as

$$\begin{aligned}
(p_0 - 2mc)\Psi - (\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})(X^{\mathrm{I}}\Psi) - (\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})(X^{\mathrm{II}}\Psi) + M(Y\Psi) &= 0 \\
- (\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})\Psi + p_0(X^{\mathrm{I}}\Psi) + M(X^{\mathrm{II}}\Psi) - (\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})(Y\Psi) &= 0 \\
- (\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})\Psi + M(X^{\mathrm{I}}\Psi) + p_0(X^{\mathrm{II}}\Psi) - (\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})(Y\Psi) &= 0 \\
M\Psi - (\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})(X^{\mathrm{I}}\Psi) - (\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})(X^{\mathrm{II}}\Psi) + (p_0 + 2mc)(Y\Psi) &= 0
\end{aligned} \tag{14}$$

where

$$M = \frac{e^2}{2c}\left(\frac{\sigma^{\mathrm{I}}\sigma^{\mathrm{II}}}{r} + \frac{(\sigma^{\mathrm{I}}\boldsymbol{r})(\sigma^{\mathrm{II}}\boldsymbol{r})}{r}\right).$$

If we omit M in (14) we obtain the result of rewriting (10). The perturbation due to the last term in (1) is thus the same as the perturbation due to M in (14) i.e.,

$$(\Delta E)_{n,n} = -c\int \{\Psi^*(Y\Psi) + (X^{\mathrm{I}}\Psi)^*\cdot(X^{\mathrm{II}}\Psi) + (X^{\mathrm{II}}\Psi)^*(X^{\mathrm{I}}\Psi) + (Y\Psi)^*\Psi\}d\tau$$

the integration being extended over the configuration space. Omitting M we obtain sufficiently exact expressions for $(X^{\mathrm{I}}\Psi)$, $(X^{\mathrm{II}}\Psi)$, $(Y\Psi)$ to substitute into this integral

$$\begin{aligned}
(X^{\mathrm{I}}\Psi) &\cong \frac{1}{2mc}(\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})\Psi,\ (X^{\mathrm{II}}\Psi) \cong \frac{1}{2mc}(\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})\Psi, \\
(Y\Psi) &\cong (\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})(\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})/4m^2c^2.
\end{aligned}$$

Substituting we obtain

$$\begin{aligned}
(\Delta E)_{n,n} = \frac{1}{4m^2c}\int \Psi^*[&(\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})M(\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}}) + (\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})M(\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}}) + M(\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})(\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}}) \\
&+ (\sigma^{\mathrm{I}}\boldsymbol{p}^{\mathrm{I}})(\sigma^{\mathrm{II}}\boldsymbol{p}^{\mathrm{II}})M]\Psi d\tau.
\end{aligned}$$

This agrees exactly with the first order perturbation energy of the term in $(e^2/8m^2c^3)$ of Eq. (44) of reference (1). In the approximation of that reference

the elimination of $(X^{I}\Psi)$, $(X^{II}\Psi)$, $(Y\Psi)$ from (14) with M omitted leads to all the remaining terms of that Eq. (44) with the exception of the term $-(e^4/16mc^3)X^2$. In the calculation of the energies of He we may use therefore Eq. (48) provided the e^4 are omitted. According to the calculations previously made[3] the results are in as good agreement with experiment as could be expected.

According to the view expressed above we may expect that the approximation of equation (10) corrected by (12) may have a wider range of application than the He fine structure. Two K electrons around a bare nucleus would be expected to obey this equation. The experimentally accessible K lines are unfortunately complicated by the screening due to other electrons.

In comparison with the calculation of reference 2 the present treatment has the advantage of being guided more closely by physical considerations in the choice of approximations. These are seen clearly only in the 16 component form of Eq. (1). The reduction to four "large" components[1] obliterated the direct significance of terms as "expectations" (Erwartungnwerte). The use of the form (14) makes it possible to carry through the applications without losing sight of the direct physical significance of the various terms.

INVESTIGATIONS IN THE FIELD OF THE ULTRA-SHORT ELECTROMAGNETIC WAVES

I. THE GENERATOR FOR THE PRODUCTION OF ULTRA-SHORT UNDAMPED WAVES

By G. Potapenko*

Norman Bridge Laboratory of Physics, California Institute of Technology

(Received November 23, 1931)

Abstract

A description of apparatus for the production of ultra-short undamped electromagnetic waves by the method of Barkhausen and Kurz is given. An investigation has been made of the method of detecting the oscillations by observing the current in the plate circuit of the generator. At a constant plate potential the current is approximately proportional to the amplitude of the oscillations. A comparison of generators with one and with two tubes shows the advantages of the former. In certain cases the energy of oscillations produced by generators with one tube can be considerably increased by a suitable choice of the "ballast" capacity of the generator.

§1. Introduction

ULTRA-SHORT electromagnetic waves, i.e., waves of the order of a few meters and shorter are acquiring an ever increasing importance in the study of electric and magnetic properties of substances and of their molecular structure. Until quite recently the above investigations were carried out by means of damped oscillations. More reliable results can be obtained using undamped waves.

Methods of obtaining ultra-short waves have been discussed in a considerable number of recently published papers.[1]

However the production of waves below 30 cm presents considerable difficulties even at the present time and waves of the order of 10 cm can be obtained only in exceptional cases. Progress is made slowly, due to considerable technical difficulties and to lack of a completely satisfactory theory. All theories suggested up to date, deal mostly with simplified cases and fail to explain the most important details. There is a fundamental difference between the theories of production of long waves and the theory of generation of short waves. In the latter case one has to take into consideration the final velocities of electrons inside of a vacuum tube. Beginning with frequencies of the order of 10^8 per sec., i.e., with waves of the order of 3 m, the period of generated oscillations becomes comparable to the time necessary for an electron to move from one electrode to another. Since the electrons move in the non-uniform field existing inside the tube, the theory presents considerable mathematical

* International Research Fellow.

[1] See for summaries: H. E. Hollmann, Zeits. f. Hochfr. **33**, 27–30, 66–74, 101–107 (1929); **35**, 21–27, 76–80 (1930); K. Kohl, Erg. d. ex. Naturwiss. **9**, 275–341 (1930).

difficulties. No complete theoretical treatment of the subject is as yet available.

In trying to obtain ultra-short waves we had to keep in mind the shortcomings of the theory of the subject and to take into account experimental data not covered by the present theory. In this respect considerable use was made of our earlier experiments.[2]

§2. Methods of Production of Ultra-Short Undamped Electromagnetic Waves

Undamped electromagnetic waves down to 1 meter can be obtained comparatively easily, using the method of "feed back coupling" well known in radio practice. Using this method K. Kohl[3] succeeded in obtaining waves 60 cm long. He worked with vacuum tubes of a special design, the whole oscillating circuit being placed within the tube.

The internal capacity of the tube and the self-inductance of the leads put a limit to a further shortening of the waves. If this method is used the only way of obtaining shorter waves is by separating overtones. C. Gutton and E. Pierret[4] obtained in this way waves 21 cm long. The low energy of the overtones and the extreme difficulty, almost impossibility, of separating them in a pure state made it necessary to develop new methods.

In 1920 H. Barkhausen and K. Kurz[5] and independently S. Zilitinkiewitch[6] proposed a new method of obtaining short waves. The method is essentially different from the method of "feed back coupling". Use is made of the electronic oscillations produced when large positive potentials are impressed on the grid with respect to the filament and the plate. Then, under certain conditions, the electrons issuing from the filament will oscillate about the grid as a position of equilibrium.

Assuming for the sake of simplicity plane electrodes H. Barkhausen and K. Kurz computed the frequency of electronic oscillations and the corresponding wave-length λ. It was found to be

$$\lambda = \frac{2000}{(E_g)^{1/2}} \cdot \frac{r_a E_g - r_g E_a}{E_g - E_a} \tag{1}$$

where E_g and E_a are the potentials (in volts) of the grid and the plate with respect to the filament; r_g and r_a are the distances of the grid and plate from the filament. For the case of cylindrical electrodes r_g and r_a will be the radii (in cm) of the grid and the plate.

In case the plate is connected directly to the filament, the potential of the plate can be taken equal to zero. Formula (1) can then be simplified as follows

[2] G. Potapenko, Zeits. f. techn. Physik **10,** 542–548 (1929).

[3] K. Kohl, Phys. Zeits. **28,** 732–734 (1927); Ann. d. Physik 1–62, (1928).

[4] G. Gutton et E. Pierret, J. de Phys. et le Rad. **7,** 15s–16s (1926).

[5] H. Barkhausen u. K. Kurz, Phys. Zeits. **21,** 1–6 (1920).

[6] S. Zilitinkewitsch, Drahtl. Tel. u. Tel. (russ) **18,** 2–22 (1923); **19,** 166–175 (1923); Arch. f. Electrot. **15,** 470–484 (1926).

$$\lambda^2 E_g = d_a^2 10^6 \tag{2}$$

where d_a is the diameter of the cylindrical plate or, since for any given tube the right side of the equation is a constant

$$\lambda^2 E_g = \text{Const.} \tag{3}$$

This simple relationship is often called Barkhausen's equation.

In their early experiments made with tubes of the usual type Barkhausen and Kurz obtained waves as short as 43 cm. It was fairly easy to move to a region of still shorter waves. At present the only other successful method of obtaining short waves is by means of magnetic fields which is being developed during recent years.[7]

S. Zilitinkiewitch[6] and later A. Scheibe[8] gave more accurate formulae for the wave-length of the oscillations, taking into account the cylindrical shape of the electrodes.

Since at first we shall be interested only in the order of magnitude, we shall use the simpler formulae (1) – (3).

These formulae indicate two possible ways of shortening the wave-length of the generated oscillations. One is by decreasing the dimensions of the tube, i.e., the radii of the grid and the plate. The other—by increasing the positive potential of the grid.

The first method does not lead us very far. In the first place, in decreasing the dimensions of the tube a natural limit is soon reached. In the second place, even before this natural limit is reached the tube ceases to produce oscillations for some, as yet, unknown reason. The second, apparently simpler, method does not lead us very far either. As the grid potential is increased more heating current must be supplied to the filament to maintain the oscillations. F. Tank[9] has shown that there exists the following relationship between the limiting length of the waves produced by the tube and the emission current required

$$\lambda^3 I_e = \text{Const.} \tag{4}$$

It is seen from the Eq. (4) that a decrease in wave-length calls for a very considerable increase in the emission current, i.e., the heating of the filament. As the heating cannot, of course, be increased indefinitely, the wave-length cannot be shortened beyond a certain limit. Thus A. Scheibe[8] as a result of his investigation, came to the conclusion that waves of 30 cm were the shortest obtainable by the method of Barkhausen-Kurz.

However, this conclusion could not be considered as final. Several cases were known, at the time we began our investigation, where the wave-length generated was found to be considerably shorter than predicted by Eq. (2).

[7] See, for example. A. Slutzkin and D. Steinberg, Ber. d. russ. phys. Ges. (russ) **58,** 395–409 (1926); Ann. d. Physik **1,** 658–670 (1929). A. Zacek, Zeits. f. Hochf. **32,** 172, (1928). K. Okabe, Proc. I. R. E. **17,** 651–659 (1929); **18,** 1748–1749 (1930).

[8] A. Scheibe, Ann. d. Physik **73,** 54–88 (1924).

[9] F. Tank, Arch. sc. Phys. et Natur. **6,** 320–321 (1924). F. Tank and E. Schiltknecht, Helv. Phys. Acta **I,** 110–138 (1928).

Thus, for example, Cl. Schaefer and J. Merzkirch[10] observed an unexpected transition of waves 66 cm long into waves 34 cn long. M. Grechowa[11] observed waves 22 and 18 cm long instead of 44 and 36 cm, as expected. Working with a constant grid potential, E. Pierret[12] obtained waves 42 cm and 18 cm long, depending on the length of the oscillating circuit applied to the tube. The observations of A. Scheibe[8] are of especial interest. He was systematically obtaining waves, whose length was approximately half of the theoretically computed value. These abnormally short waves appeared quite separately from the corresponding longer waves, which proved that the shorter waves could not be considered as the overtones of the longer waves.

Having in view these separate indications of the possibility of obtaining waves of a shorter length than predicted theoretically, we decided systematically to investigate conditions under which such waves appeared and then to attempt to obtain them in as pure a state as possible.

This way of obtaining short waves seemed particularly attractive as it was not necessary to increase the grid potential and the heating current. It was thus possible to operate the tube under more normal working conditions and waves of greater stability could be expected, which in itself is a fact of great importance.

§3. Description of the Apparatus

The fundamental scheme used for the production of ultra short waves is shown in Fig. 1. A vacuum tube V and a "ballast" condenser C of equal capacity were placed between two oscillating circuits: the plate circuit K_a and the grid circuit K_g. The circuits were formed of two copper wires 4.5 mm in

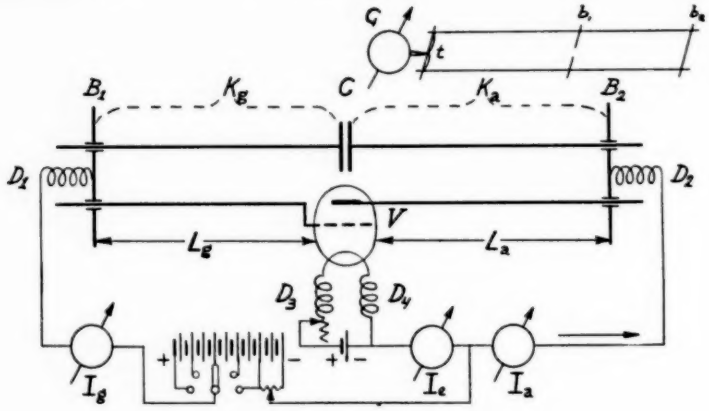


Fig. 1. Schematic diagram of the apparatus.

diameter, with two movable plate bridges B_1 and B_2. The distance between the axes of the wires was 2 cm. The lengths L_a and L_g of the circuits K_a and K_g could be varied from 5 cm to 75 cm by moving the bridges. The oscillating

[10] Cl. Schaefer u. J. Merzkirch, Zeits. f. Physik **13**, 177–178 (1923).
[11] M. Grechowa, Zeits. f. Physik **38**, 621–634 (1926).
[12] E. Pierret, C. R. **184**, 1428–1430 (1927).

circuits were shielded from the rest of the system by means of chokes D_1, D_2, D_3, D_4.

As seen from the diagram the plate and the filament are connected directly and the potential of the plate was always equal to the potential of the negative end of the filament. The grid voltage E_g was varied from 0 to +720 volts either by means of a d.c. generator and a potentiometer or by means of a battery (400 volts), as shown on the diagram, a switch and small potentiometer P. The latter method is preferable; no high resistance potentiometers are required which when connected in parallel with the grid-plate interval usually leads to errors in the determination of the grid potentials.

As to the various other details, a correct choice of chokes D_3 and D_4 placed in the heating circuit of the tube is of particular importance.

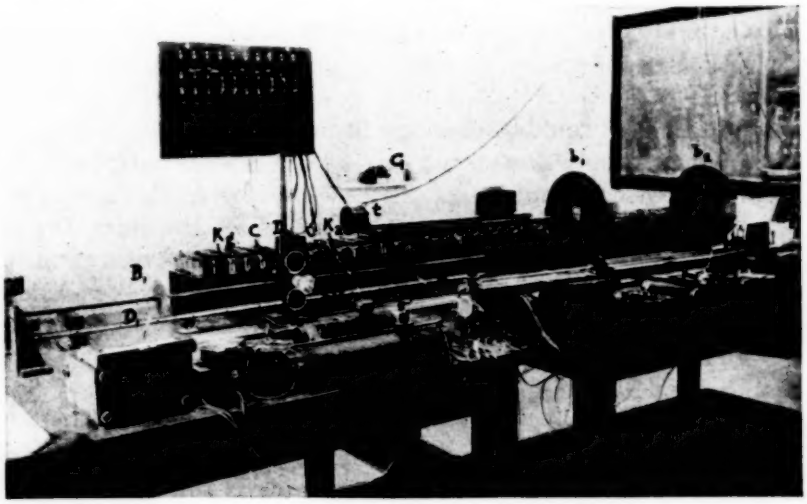


Fig. 2. General view of the apparatus.

A heating current of the order of one ampere passes through these chokes. To avoid the undesirable influence of the magnetic field of this current, it is convenient to make these chokes of two oppositely wound spirals.[13] As chokes D_3, D_4 had a fundamental wave-length,[14] app. 65 cm, additional chokes of a larger size were placed in the heating circuit when working with waves longer than 50 cm. Chokes D_1 and D_2 are straight spirals of the same dimensions as chokes D_3 and D_4. They were connected to the outer side of bridges B_1 and B_2, midway between the wires of the plate and the grid circuits.

The wave-length of the generated oscillations was measured by means of a Lecher system, a thermocouple t and mirror galvanometer G (see Fig. 1). The thermocouple was connected either directly to the bridge of the Lecher

[13] R. Gunn, Proc. I.R.E. **15**, 801–808 (1927).

[14] Each of the spirals used consisted of 14 turns of a copper wire 1 mm in diameter. Radius $r_1=3$ cm, $r_2=3$ mm. When so constructed, the self-inductance and the capacity of the choke were $L=0.36\times10^{-6}H$ and $C=0.66\times10^{-12}F$ thus giving equal capacitive and inductive resistances for $\lambda=30$ cm.

system,[15] or was placed in an aperiodic circuit, loosely coupled with the Lecher system.

The Lecher system was loosely coupled with the plate circuit. Coupling with the grid circuit was established only in special, specifically mentioned cases.

§4.

It is very important to maintain the heating current of the vacuum tube constant when working with ultra-short waves. When measurements are made at a constant grid voltage a very sensitive and convenient method of controlling the heating current is by means of the emission current. Unfortunately, this method is not applicable when the grid voltage E_g varies. It is known that even within the region of saturation current, the emission current varies appreciably with E_g. Therefore, the steadiness of the emission current with varying grid voltage would indicate variations in the heating current.[16] In our experiments the grid voltage was constantly varying. The heating current was controlled directly by means of an accurate ammeter. The disintegration of the filament could be observed through a change in its resistance or through a change in the emission current at some definite fixed point.[17]

Together with the steadiness of the heating current another important feature is the steadiness of the internal "regime" of the tube. It depends largely on the temperature of the tube and on its previous operation. It was found that even a short overheating, resulting in a barely observable disintegration of the filament, has a very great effect on the work of the tube. It may happen that observations made after the overheating will be found in no way comparable with the earlier observations. If observations have to be interrupted for some time, it is found best to "preheat" the tube before beginning a new series of measurements, i.e., to start the heating current some 20–40 minutes before the beginning of measurements. By this time the temperature of the tube comes to a steady value and measurements give comparable results.

§5.

A large number of vacuum tubes have been tested. The following were found most suitable for the production of ultra-short waves: Tubes of the type *R* 5 and *J* 4 (Russian); of the type *TKD-VT* 49 (German); of the type Phillips *E* (Dutch); type *MT* 5 (English); and tubes of the type Metal *TMC* and Metal *E* 4*M* (French). So far we have not been able to obtain good results with any of the American tubes which we had at our disposal.

All the above mentioned tubes have pure tungsten filaments and cylindrical grids and plates. We could observe no ultra-short waves with tubes having plane electrodes or other kinds of filaments.

[15] G. Potapenko, Trans. Phys. Res. Inst. Univ. of Moscow (russ) **6,** 1–103 (1926)—sec. 6.

[16] The importance of complete clearness on this point is seen from a recent discussion: M. J. O. Struth, Ann. d. Physik **4,** 17–32 (1930); **8,** 794–796 (1931); H. E. Hollmann, Ann. d. Physik **5,** 247–260 (1930).

[17] We used as such a point $E_g = 220$ volts and $L = 50$ cm in which no oscillations were produced by the tubes *R*5.

§6. Detection of Oscillations

The selection of a method of detecting oscillations and of measuring their intensity is one of the most essential questions in all work with ultra-short waves.

In working with a scheme like ours the simplest method of detecting oscillations is by observing the plate current. If the potential of the plate is zero with respect to the negative end of the filament (as it is the case in our scheme), no plate current can appear unless oscillations be present. Its direction will correspond to the flow of electrons toward the plate. Its strength must correspond to the energy or more exactly (as we shall see later) to the amplitude of the oscillations of the plate potential. The plate current is frequently taken as a direct measure of the energy of the oscillations. This assumption is based on the fact that plate current curves are similar to the

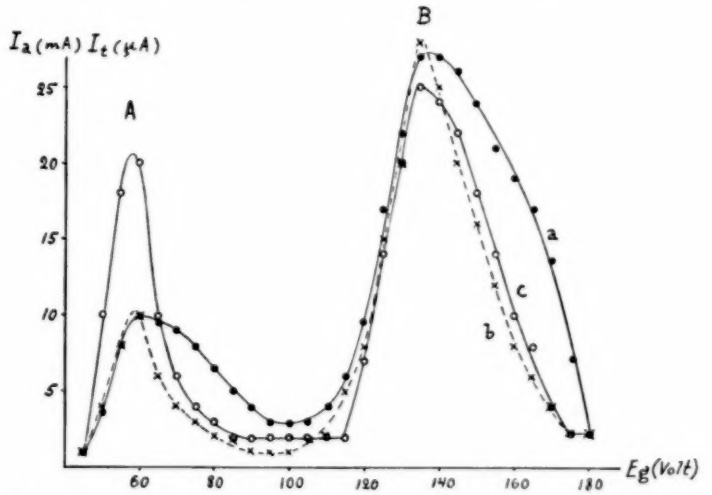


Fig. 3. The curves of the plate current and thermocouple current at $L=60$ cm; heating current $I_h=0.70$ A. (*a*)—plate current; (*b*)—thermocouple current; aperiodic circuit at B_1; (*c*)—thermocouple current; aperiodic circuit at the middle of circuit K_a.

curves of the current of a thermocouple placed in the oscillating circuit or in one of the auxiliary aperiodic circuits coupled with the oscillating circuit.[18]

In the preliminary experiments simultaneous measurements were made of the plate current and of the energy of oscillations. The latter was determined by means of an aperiodic circuit with a thermocouple. The aperiodic circuit was placed near the bridge B_1 or B_2 (see Fig. 1). Thus it was loosely coupled with the plate or the grid circuits of the tube.

Fig. 3 gives the results of observations. Plate current and thermocouple currents are plotted against the grid voltage at $L_a=L_g=L=60$ cm. There is a similarity between the two curves (curves *a* and *b*).

Their maxima correspond to the same values of the grid voltage. However, there is no strict proportionality between the two curves. After passing

[18] See for example J. Tank and E. Schiltknecht, reference 9, p. 127.

through a maximum the energy of oscillations declines more sharply than the plate current. Analogous results were obtained in all other measurements, with different values of L_a and L_g.

To determine more accurately the relation between the plate current and the thermo current we investigated their dependence on the heating current. A change in the heating current will affect the amplitude of oscillations, hence the plate and thermo currents will also be affected.

Table I gives the results of such measurements, made for two values of L_a and L_g. The grid voltage E_g was maintained at a value corresponding to a maximum of the plate current, which, as we have seen, approximately corresponds to the maximum of the thermo current.

TABLE I. *Tube R5 (No. 20).*

$L_a = L_g$ (cm)	I_h (amp.)	E_g (volt)	I_a (mA)	I_t (mA)	I_t/I_a	I_t/I_a^2
30	0.700	100	2.5	30.0	12.0	4.8
30	0.690	106	2.2	24.0	11.0	5.0
30	0.680	115	1.85	18.8	10.2	5.5
30	0.670	123	1.3	12.0	9.2	7.1
30	0.660	130	1.0	7.9	7.9	7.9
30	0.650	128	0.35	1.0	3.0	8.3
24	0.700	142	2.5	24.0	9.6	3.8
24	0.690	152	2.2	18.8	8.5	3.9
24	0.680	153	1.65	12.8	7.7	4.2
24	0.670	158	1.15	7.5	6.5	5.7
24	0.660	163	0.45	1.2	2.7	6.0

Table I shows that a change in the heating current produces similar changes in the plate and thermo currents.

The last two columns of Table I give the values of the ratios I_t/I_a and I_t/I_a^2. It is seen that the ratios vary with the heating current. The two columns vary in opposite directions. However I_t/I_a varies about three times faster than I_t/I_a^2. Thus, strictly speaking, the plate current is proportional neither to the energy of the oscillations, nor to their amplitude. It corresponds, however, more closely to the amplitude of the oscillations. The above measurements were made at a constant plate voltage. So we may say: *when the plate voltage is kept constant, the plate current may be taken as approximately proportional to the amplitude of oscillations.* In our case the plate was connected directly to the filament and the above restriction was automatically satisfied. It is clear that when the plate voltage varies we cannot expect any correspondence between the plate current and the amplitude of the oscillations.[19] It is obvious that we shall have plate current at any positive plate potential, even in the absence of oscillations. Conversely, there will be no plate current of the direction we consider, when the plate has a sufficiently large negative potential, i.e., when its absolute magnitude will be greater than the amplitude of the oscillations.

[19] Comp. W. Kroebel, Zeits. f. Physik **61**, 242 (1930).

§7.

The plate current is only an approximate measure of the energy of oscillations. It is very convenient to use, however, as it is hundreds, even thousands of times, greater than the current from the thermocouple. In our measurements we used the plate current. The thermo current was measured only when the plate potential had to be changed.

Erroneous results are frequently obtained when working with an aperiodic circuit and a thermocouple. If the aperiodic circuit be removed from the bridge and brought near the middle of the oscillating circuit, instead of the curve *b* we obtain curve *c* Fig. 3 giving an entirely different idea of the relative intensity of oscillations in regions *A* and *B*. This is due to the fact that oscillations *A* and *B* had a different wave-length ($\lambda \sim 100$ cm and $\lambda \sim 60$ cm). Moving the aperiodic circuit brings it near the loop of the first wave and the node of the second one. The maximum of the first wave was thus increased. Hence, to obtain correct results when measuring the energy of waves of different length, the aperiodic circuit must be kept in a fixed position with respect to the nodes and loop of the waves.

In our experiments the aperiodic circuit was placed near one of the bridges. It would seem to be preferable to place the thermocouple (or detector) directly in the bridge, as was done by E. Gill and J. Morrell[20] and H. Hollmann.[21] However, the sensitivity of this method depends greatly on the resistance (capacitance) of the bridge, which may vary considerably with the wave-length as we are near one of its fundamental periods when working with waves within the range of 10–100 cm.

The grid current I_g can also be used for measuring the intensity of oscillations. We have

$$I_a + I_g = I_e. \tag{5}$$

The emission current I_e varies little within the saturation region (where we worked) and at constant heating current. From the Eq. (5) it is seen that as the plate current varies, the grid current must vary in an opposite direction. The plate current, however, seldom exceeds 15 percent of the grid current. Variations in the grid current are much harder to observe.

When the intensity of oscillations is low, it becomes extremely difficult to observe variations in the grid current. In our work, measurements of the grid current were made only in very rare cases.

§8. Generator with One and Two Tubes

In a scheme like ours, where the tube *V* is paired with a condenser *C* of equal capacity (see Fig. 1) it would seem natural to replace the condenser with another vacuum tube and to operate the two tubes simultaneously. The resulting scheme would be equivalent to the one proposed by F. Holborn[22] for the production of short waves by the method of feed back coupling. M. T.

[20] E. Gill and L. Morrell, Phil. Mag. **44,** 161–178 (1922).
[21] H. E. Hollmann, Ann. d. Physik **86,** 129–188 (1928).
[22] F. Holborn, Zeits. f. Physik **6,** 328–338 (1921).

Grechowa[23] and E. Pierret[24] used this scheme for the production of short waves by the method of Barkhausen-Kurz. The latter found that: (1) in the presence of a second tube the oscillatory energy is more than doubled[25] and (2) if there is a slight difference in the length of waves produced by each tube when working separately, they will produce, when working together, waves of only one length. The wave-length will be shorter than that produced by each tube separately.

In the beginning of our investigation we tried to use this scheme. We soon discovered the difficulty of finding two tubes with similar properties. Several hundreds of tubes of the type *R*5 were tested and only 4 or 5 were found

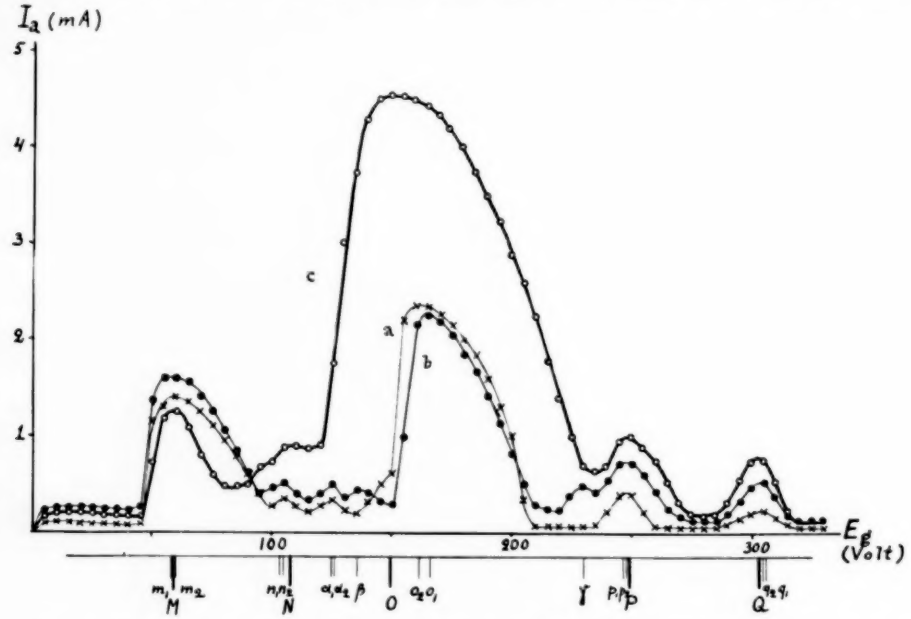


Fig. 4. (I_a, E_g)-Characteristics at $L=50$ cm; $I_h=0.70$ A. (a)—Vacuum tube No. 20, "ballast" tube No. 5. (b)—Vacuum tube No. 5, "ballast" tube No. 20. (c)—Vacuum tube No. 20 and No. 5 working simultaneously.

whose characteristics were similar with respect to the number of regions of oscillations, their location and the energy of oscillations.

Fig. 4 shows the relation between the plate current and the grid voltage for two such tubes (Nos. 5 and 20). When one tube was lighted the other was used merely as a condenser C (see Fig. 1). Such curves will be called the (I_a, E_g)-characteristics.

The lengths of the plate and grid circuits and the heating current were always kept constant, while these curves were obtained. Fig. 4 shows also the (I_a, E_g)-characteristic for the two tubes working simultaneously. It is seen

[23] M. T. Grechowa, Trans. Electr. Res. Inst. (russ) **11**, 10–18 (1925); Zeits. f. Physik **35**, 50–58 (1925).

[24] E. Pierret, reference 12.

[25] See also A. Scheibe, Jahrb. dr. Tel. u. Tel. **27**, 1–5 (1926).

that this characteristic is not a sum nor a definite function of the individual characteristics of the two tubes. In the regions N and O the energy of oscillations is greater than the sum of the energies of the individual oscillations n_1, n_2 and o_1, o_2 respectively. In the region M the oscillatory energy is considerably less than each of the energies m_1, m_2. Finally in the region P the total oscillatory energy barely exceeds the energy of individual oscillations P_1 and P_2.

Hence E. Pierret's first conclusion cannot be extended to all cases of such oscillations. The same may be said of his second result. With two tubes working simultaneously, the wave-length obtained was shorter or longer than the wave-length produced by one tube. When the characteristics of the two tubes differed more than those shown in Fig. 4, the resulting oscillations consisted of waves of two or more frequencies. Fig. 5 gives (I_a, E_g)-characteris-

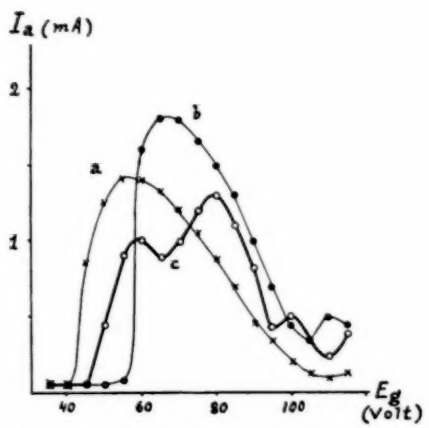


Fig. 5. (I_a, E_g)-characteristics at $L=50$ cm; $I_h=0.70$ A. (a)—Vacuum tube No. 20 "ballast" tube No. 16. (b)—Vacuum tube No. 16 "ballast" tube No. 20. (c)—Vacuum tube No. 20 and No. 16 working simultaneously.

tics for a pair of tubes differing more than tubes No. 20 and No. 5. There is no regularity of any sort in the addition of the oscillations. It is seen that the resulting energy of oscillations produced by two tubes is less than the energy produced by each tube separately. This was the case for region M on Fig. 4.

The curves of Fig. 4 show that the resulting (I_a, E_g)-characteristic for the two tubes is lacking in finer details which are seen on the characteristics of single tubes. In fact, regions α_1, α_2, β, γ are so to speak "absorbed" by region O, as is seen on Fig. 4. Preliminary work had shown, however, that in the study of oscillations these small regions of low energy are of particular interest. Accordingly, it has been decided to use only one tube for the production of oscillations. This tube was balanced against a small condenser of variable capacity, or against another tube of the same type and of equal capacity. This second tube was used as a condenser. It was not lighted during the experiments. We shall call it a "ballast" tube.

§9. The "Ballast" Capacity

The capacity of the condenser C (see Fig. 1) or of the "ballast" tube must be as nearly equal to the capacity of the working tube as possible. It is not sufficient to have a "ballast" tube of the same type as the working one. Two tubes of the same type may differ in capacity by as much as 10–15 percent. Fig. 6 shows the effect of such a difference in capacity on the operation of the generator. It shows four (I_a, E_g)-characteristics of the tube of type R5.

One of the characteristics is obtained by using as "ballast" tube, No. 18, of similar type. The second characteristic was obtained with another "ballast" tube, No. 6, of a similar type. The first characteristic is very similar to the one when tube No. 20 was paired with tube No. 5 (see Fig. 4). The second characteristic shows a total absence of oscillations in the region 0.

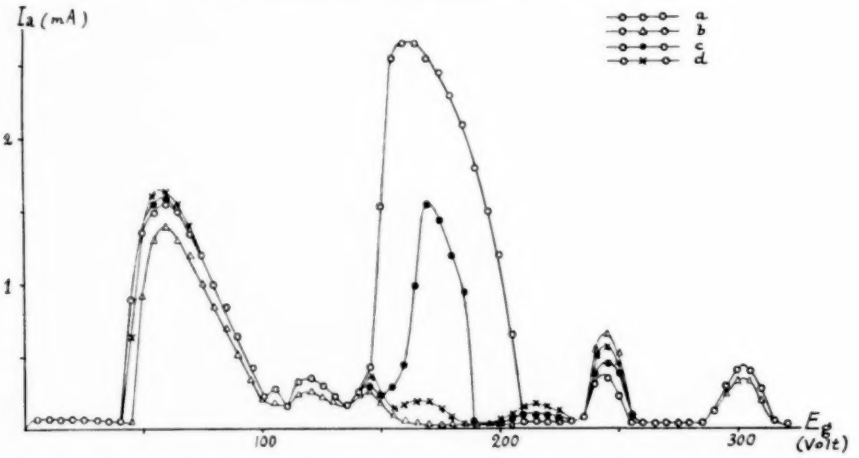


Fig. 6. (I_a, E_g)-characteristics at $L = 50$ cm; $I_\Lambda = 0.70$ A (a)—Vacuum tube No. 20, "ballast" tube No. 18. (b)—Vacuum tube No. 20, "ballast" tube No. 6. (c)—Vacuum tube No. 20, "ballast" tube No. 18 + Capacity 0.12 cm. (d)—Vacuum tube No. 20, "ballast" tube No. 18 + Capacity 0.18 cm.

It was found that tube No. 6 had an internal capacity approximately 0.2 cm greater than the other tubes.[26]

To prove that the change of characteristics was really due to the difference in capacity, a small condenser of variable capacity was connected between the plate and grid circuits at the tube No. 18, in parallel with its plate-grid capacity.

Two of the characteristics of tube No. 20 taken with this condenser are shown in Fig. 6. The first one, taken with the condenser capacity $C_1 = 0.12$ cm, shows a considerable decrease of oscillatory energy in the region 0. The second characteristic, taken with the condenser capacity $C_2 = 0.18$ cm, shows almost a complete absence of oscillations in region 0. The characteristic is very similar to that of the pair of tubes No. 20–6. At the same time we

[26] The capacity of tubes Nos. 5, 18, 20 measured at $\lambda = 1$ m was 2.85 – 2.9 cm, the capacity of tube No. 6, 3.1 cm.

observe an increase in the intensity of oscillations in the region P, and no changes in regions M and Q.

This difference in the effect of the presence of the condenser on different regions can be explained from a cousideration of the distribution of the loops and nodes of the standing waves in the plate and the grid circuit. Fig. 7(a) shows the distribution of potential for the region M, Fig. 7(b) and 7(c) for the region 0. It is seen that when the loops are not situated in the immediate neighborhood of the tubes (region M) a small difference in their capacity has a small effect on the characteristic.

When, on the other hand, the loops are in the neighborhood of the tubes (region 0) a difference in capacities has a very pronounced effect on the standing waves and the energy of the oscillations falls off sharply.[27]

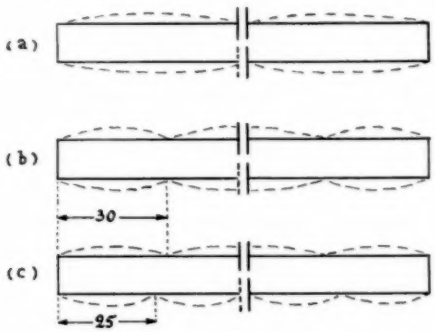


Fig. 7. Distribution of the potentials in the plate and the grid circuits. Vacuum tube No. 20 at $L = 50$ cm. (a)—Region M, ballast tube No. 18 or No. 16. (b)—Region O, ballast tube No. 18. (c)—Region O, ballast tube No. 18 with condenser 0.18 cm or ballast tube No. 6 without condenser.

The above considerations show the importance of having the capacity of the "ballast" tube equal to that of the working tube.

Obviously, the working and the "ballast" tubes must have sockets of equal capacity. This can easily be checked by interchanging the positions of the working and the "ballast" tube. If the capacities of the sockets are equal such an interchange will produce only an insignificant change in the characteristic.

In conclusion the author wishes to express his thanks to the Rockefeller Foundation for the grant of a Fellowship and to Professor R. A. Millikan for the facilities of the Norman Bridge Laboratory.

[27] Under such circumstances oscillations are produced such that a second wave train is superimposed on the one already present. For this second train $\lambda^2 E_g$ is below the usual value. The above refers to oscillations which in future articles we shall call normal oscillations.

INVESTIGATIONS IN THE FIELD OF THE ULTRA-SHORT ELECTROMAGNETIC WAVES

II. THE NORMAL WAVES AND THE DWARF WAVES

By G. Potapenko*

Norman Bridge Laboratory of Physics, California Institute of Technology

(Received November 23, 1931)

Abstract

The results are presented of an investigation of the production of ultra-short undamped electromagnetic waves by using the method of H. Barkhausen and K. Kurz.

Method of working diagrams. Normal waves and dwarf waves. A method is developed for the graphic representation of the work of generators of ultra-short waves. This method is based on the construction of special "working diagrams." These diagrams define the location of "regions of oscillations," which show the values of the natural periods of the oscillating circuits and the values of the grid potentials at which oscillations are generated. Vacuum tubes can generate two kinds of ultra-short waves. The first kind have a wave-length approximating that computed by Barkhausen's formula $\lambda^2 E_g = d_a{}^2 10^6$. Their period is nearly equal to the time required for the electrons to move from the filament to the plate and back (normal waves). The second kind of waves are considerably shorter (dwarf waves). Both kinds of waves satisfy the equation $\lambda^2 E_g = \text{const.}$ for points on the working diagram where the plate current (the amplitude of the oscillations) has its maximum value.

Complex working diagrams. Dwarf waves of higher orders. Vacuum tubes can have complex working diagrams with a large number of regions of oscillations. In such a case the tube generates different dwarf waves. Their length is two, three and four times shorter than that of the normal waves. Dwarf waves are accordingly divided into waves of the 1st, 2nd, 3rd, etc. orders. The shortest dwarf waves of the 4th order, generated by tubes of the type *R*5, had a wave-length $\lambda = 9.4$ cm. The presence of dwarf waves of higher orders shows that *vacuum tubes can generate oscillations of a frequency considerably greater than the frequency of the electronic oscillations.* Both the normal and dwarf waves belong to the same type of GM-oscillations. Limits were determined within which Barkhausen's formula is applicable. It is shown that the difference in the number of regions of oscillations on the working diagrams depends on the difference in the time required for the electrons to pass in different directions within the tube. The latter depends on the asymmetry in the arrangement of the electrodes.

The nature of dwarf waves. Dwarf waves are oscillations of the circuits within the tube or coupled with the tube which are excited in such a manner that during the time τ it takes for the electrons to pass from the filament to the plate and back, the circuits perform two complete oscillations (dwarf waves of the 1st order), three complete oscillations (dwarf waves of the 2nd order) etc. Thus the wave-lengths are equal to: $\lambda_0 = c_0\tau$ (normal waves), $\lambda_1 = c_0\tau/2$ (dwarf waves of the 1st order), $\lambda_2 = c_0\tau/3$ (dwarf waves of the 2nd order), $\lambda_3 = c_0\tau/4$ (dwarf waves of the 3rd order), etc. Dwarf waves 9.5 − 18.5 cm long originate in oscillating circuits, which are inside the tube. The advantages of dwarf waves of higher orders are shown, owing to the possibility of using lower grid potentials, which leads to a greater steadiness in the operation of the tube.

* International Research Fellow.

§1. Introduction

THE present article gives some of the results of an investigation on the production of ultra-short undamped electromagnetic waves. These waves were obtained using the method of H. Barkhausen and K. Kurz by means of a generator which has been described earlier.[1]

§2. The Method of Working Diagrams. Normal Waves and Dwarf Waves

There are two schemes of graphical representation of the work of vacuum tubes generating ultra-short waves by the method of H. Barkhausen and K. Kurz. Both schemes can be used regardless of what oscillating circuit the tube is connected with or of what generator it makes a part. The first method is to represent the work of a tube (or generator) by means of curves showing the relation between the energy of oscillations (or the direct current I_a in the plate circuit of the tube) and the grid potential E_g. The dimensions of the oscillating circuits connected with the tube, their natural periods and the heating current are kept constant. We have already used such curves (I, §8), and called them the (I_a, E_g)-*characteristics* of the generator. The second scheme consists in representing the energy of oscillations or the plate current as a function of the dimensions or (what is the same thing) of the natural periods of the oscillating circuits. The grid voltage and the heating current are kept constant. We can call such curves the (I_a, L)-*characteristics*, where L stands for the dimensions or for the natural periods of the oscillating circuits.

As a matter of fact, the energy of the generated oscillations depends both on the electrode potentials and on the natural periods of the oscillating circuits. Thus, neither of the characteristics gives a complete picture of the work of the tube.

This drawback can be obviated by systematically investigating the work of a tube, varying simultaneously the potentials and the periods of the oscillating circuits. This is easily accomplished. The results can be conveniently and clearly presented by means of special "working diagrams," which we shall presently discuss. This "*method of working diagrams*," as it may be called, has a number of advantages and permits us to give complete characteristics of short wave generators. This method does not preclude the possibility of using the (I_a, E_g)-and (I_a, L)-characteristics. The latter are useful in representing separate details of the working diagrams.

§3.

We shall now investigate the work of the generator described above, using tubes of the type *R*5. This is a tube commonly used in Russia for radio reception. It shows considerable steadiness in operation, not only at normal heating (I_h=0.68A; I_a=5–7 mA) but even at considerable overheating (I_a= 20–30 mA and higher).

Table I gives the results of measurements of the dependence of the plate

[1] G. Potapenko, Phys. Rev. **39**, 625 (1932); this article will be referred to as I.

TABLE I. E_g (*volt*). *R5* (*No. 2+No. 4*) $I_k = 1.36A$.

300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
280	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
260	0	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	0	0.5	1.2	0	0	0	0	0	0	0	1.8	1.3	0	0	0	0	0	0	0.2	0.1	1.2	0	0	0
220	0	0.3	2.0	2.6	0.9	0	0	0	0	0	0.4	2.6	2.2	1.2	0	0	0	0	0.2	0.2	2.6	2.3	1.7	0.8
200	0	0	0.4	3.5	2.6	1.7	0.6	0	0	0	0.2	3.2	3.1	2.8	2.1	1.1	0	0	0	0.1	2.9	3.2	3.0	2.8
180	0	0	0.1	4.0	3.7	3.2	2.4	1.1	0.1	0	0	0.5	3.2	3.4	3.2	2.8	2.0	1.6	0	0	0.2	0.5	3.5	3.7
160	0	0	0	2.3	4.1	3.9	3.6	2.8	1.7	0.1	0	0	0.4	2.1	3.7	3.7	3.2	3.1	2.1	0.4	0	0.1	0.5	3.9
140	0	0.2	0	0.3	2.3	3.6	4.1	4.0	3.5	2.3	0.7	0	0	0.1	0.5	2.4	3.9	3.9	3.6	3.3	2.3	1.7	0.6	0.5
120	0	0	0.9	0.1	0	0.4	2.1	4.1	4.1	3.8	2.8	1.5	0.2	0	0	0	0.5	2.0	3.8	4.0	3.7	3.4	2.8	2.2
100	0	0	0	1.4	1.6	0.1	0	1.8	3.6	4.2	4.0	3.4	2.3	1.4	0.4	0	0	0	1.9	2.0	3.8	3.9	3.9	3.8
80	0.7	0.2	0.3	0	0	0	0	0	0	2.8	4.0	4.0	3.7	3.2	2.3	1.6	0.9	0.5	0	0	0	0.1	3.3	3.8
60	1.8	1.2	0.7	0.3	0.2	0.1	0.1	0	0	0	0.1	0.1	0.4	3.4	3.2	3.0	2.5	2.1	1.5	1.3	0.9	0.5	0.3	0.2
40	1.6	1.0	0.6	0.6	0.6	0.5	0.4	0.3	0.1	0	0	0	0	0.1	0.2	0.2	2.0	1.8	1.6	1.4	1.3	1.2	1.1	1.0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74 L.(cm)

current I_a on the grid voltage E_g and on the lengths of the plate and the grid circuits $L = L_a = L_g$ (see I, Fig. 1). These measurements were made at the beginning of our investigation with two tubes Nos. 2 and 4 working simultaneously.[2]

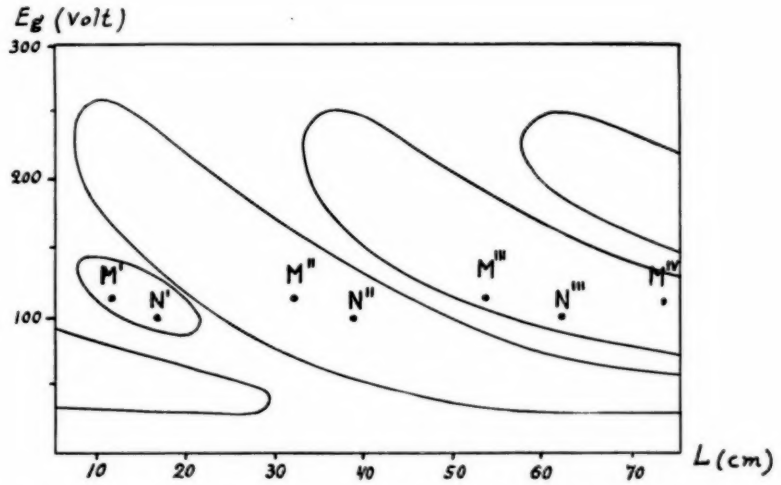


Fig. 1. A working diagram for vacuum tubes of the type *R*5. The regions of oscillation show the length of the oscillating circuits and the grid potentials at which oscillations are generated.

In our previous paper it was shown (I, §6) that the presence of the plate current indicates the presence of oscillations. The current is approximately proportional to the amplitude of the oscillations. Table I shows that the plate current and therefore the oscillations exist at definite values of E_g and L, ly-

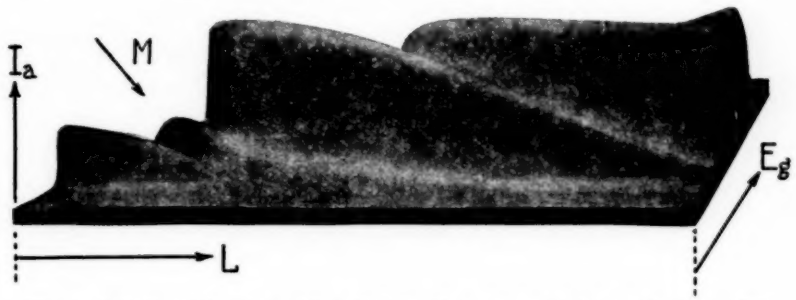


Fig. 2. A 3-dimensional working diagram for vacuum tubes of the type *R*5. The ridge *M* corresponds to dwarf waves.

ing within certain regions, which can be called "*regions of oscillations.*" By plotting the locations of regions of oscillations, we obtain Fig. 1. Such diagrams will be called *working diagrams.*

To get an idea of the general appearance of the regions of oscillations, we can construct a *3-dimensional working diagram* shown in Fig. 2. It is seen from

[2] The numbers are used merely to indicate the order in which the tubes were tested.

the diagram that all regions of oscillations have the same general appearance of continuous ridges. Their height corresponds to the value of the plate current in a given region, or as we have shown, to the amplitude of the oscillations. As all the tubes investigated gave working diagrams of the same type, there is no need to construct 3-dimensional working diagrams in each case. A simple working diagram, similar to the one of Fig. 1 will be sufficient. It should be mentioned that a simple working diagram gives also an approximate idea of the relative magnitude of the amplitude of the oscillations: the size of the area of a region of oscillations usually corresponds to the magnitude of the amplitude of the oscillations. A simple comparison of diagrams Figs. 1 and 2 shows the correctness of this statement. It is evident that a working diagram will give a more complete idea of the work of a generator than separate (I_a, E_g)-or (I_a, L)-characteristics, because it is a combination of all

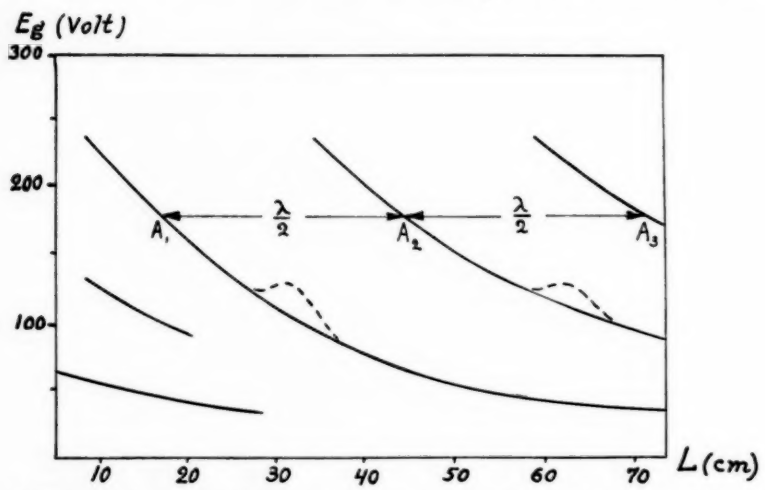


Fig. 3. Simplified working diagram for vacuum tubes of the type *R*5. —lines of maxima corresponding to the maximum values of the plate current, i.e., to the maximum values of the amplitude of the oscillations. ---bends on the lines of maxima indicating a leak of energy through the chokes.

such characteristics. It is seen that the (I_a, E_g)-characteristic represents a section of a 3-dimensional working diagram cut by a plane parallel to the E_g axis and the (I_a, L) characteristic is a section cut by a plane parallel to the L axis.

The work of measuring and of plotting the diagrams is greatly simplified if the position of only those points is determined at which the amplitude of the oscillation has a maximum value. If such points are plotted on the diagram and connected by smooth curves, a simple diagram is obtained which we shall call a *simplified working diagram*. Such a diagram is shown on Fig. 3. It is evident that the curves of such simplified diagrams or *lines of maxima* as they may be called correspond to the peaks of the ridges of the 3-dimensional diagrams. Unfortunately, the simplified diagrams do not give any idea of the relative energy of the oscillations. When such figures are of particular

interest they can be given directly on the simplified diagram or the simplified diagram can be supplemented by one or two (I_a, E_g)-or (I_a, L)-characteristics.

The working diagrams have the following important property: from their general appearance it is frequently possible to make an approximate determination of the wave-length of the generated oscillations. Consideer, for example, the diagram of Fig. 3; we draw a line $E_g = 180$ volts, intersecting all the three lines of maxima. The points of intersection A_1, A_2 and A_3 will correspond to circuit lengths $L_1 = 17$ cm, $L_2 = 44$ cm, $L_3 = 71$ cm. The distance between two consecutive points will be 27 cm. If we determine the wave-length of the oscillations at these three points, they will be found to be nearly equal and to have an average value $\lambda = 54$ cm. Hence it follows that the distance (expressed in centimeters) between two corresponding neighboring points on the diagram equals approximately half the wave-length of the oscillations at these points.[3] By *corresponding points* we mean points lying in two similar regions of oscillation and corresponding to the same value of E_g.

§4.

What is the wave-length of oscillations in different regions on the diagram?

We determine the wave-lengths in such points of the regions of oscillations where the plate current has a maximum value. These points will evidently lie on the peaks of the ridges of the 3-dimensional working diagram or on the lines of maxima of the simplified working diagram. Table II gives the

TABLE II. *R5 (No. 2+No. 4). $I_h = 1.36A$.*

E_g (Volt)	L (cm)	λ (cm)	$\lambda^2 E_g$	E_g (Volt)	L (cm)	λ (cm)	$\lambda^2 E_g$
40	65	114.0	5.20×10^5	82	21	48.9	1.96×10^5
50	5	109.0	5.94 "	95	19	46.2	2.09 "
80	40	80.0	5.12 "	100	17	44.9	2.01 "
100	71	74.4	5.53 "	110	11	42.1	1.95 "
120	30	66.0	5.23 "	120	9	40.9	2.01 "
120	62	68.8	5.68 "				
140	53	61.5	5.29 "				$(\lambda^2 E_g)_{aver.} = 1.99\times10^5$
150	50	59.7	5.34 "				
160	20	60.4	5.84 "				
160	74	56.9	5.18 "				
180	41	54.2	5.29 "				
200	65	52.5	5.51 "				
210	11	51.6	5.59 "				
			$(\lambda^2 E_g)_{aver.} = 5.44\times10^5$				

results of such measurements. It is seen from the table that there are two groups of points. Within the limits of experimental error, Barkhausen's equation

[3] This rule is evidently analogous to the well known rule concerning the distance between two regions of oscillation on the (I_a, L)—characteristics, which was first formulated by Zilitinkewitch, Drahtl. Tel. u. Tel. (russ) **18,** 2–22 (1923); Arch. f. Electr. **15,** 470–484 (1926).

$$\lambda^2 E_g = \text{Const.} \tag{1}$$

holds true for each of the groups. However, the absolute value of the product $\lambda^2 E_g$ is several times greater for one group than for the other. We have seen (I, §2) that in our case, when the plate of the tube is directly connected to the filament, the following approximate expression can be derived from theoretical considerations:

$$\lambda^2 E_g = d_a^2 10^6. \tag{2}$$

The diameter d_a of the internal surface of the plate of tube R5 is equal to 0.87 cm. Substituting this value of d_a in (2) we get $\lambda^2 E_g = 7.6 \times 10^5$. This computation is not very accurate in view of the simplifying assumptions made in deriving Eq. (2). Still, the calculated value of the product $\lambda^2 E_g$ is not very far from the value observed for the first group of points.[4] Waves corresponding to this first group of points may be called normal Barkhausen's waves or simply *normal waves*. The second group of waves having a considerably shorter wave-length will be called *dwarf waves*.

Two types of Barkhausen's oscillations are known at the present time: (1) the *BK-oscillations*, when the length of the generated waves depends on the electrode potentials only, and is independent of the lengths or of the natural periods of the oscillating circuits connected with the tube, and (2) the *GM-oscillations*, when the length of the generated waves depends on the length (or the natural period) of the oscillating circuits. Also, as it was first shown by E. Gill and J. Morrell,[5] Eq. (1) is satisfied only at voltages corresponding to a maximum of the energy of oscillation.[6] Thus, in the case of BK-oscillations Eq. (1) is satisfied always,[7] i.e., for all grid voltages and for all lengths of circuits. In the case of GM-oscillations Eq. (1) is satisfied only for points corresponding to the maximum values of the plate current.

The observations presented in Table II refer to points where the plate current was at a maximum and therefore give no information as to which of the two types the observed observations belong. To obtain such information, observations were made at points not corresponding to the maximum of the plate current. The results are given in Table III. They cover two cases: the lengths of the oscillating circuits were varied at constant grid voltgae and the opposite case when the grid voltage was varied and the length of the oscillating circuits was kept constant. In both cases observations were made at points which did not correspond to the maximum values of the plate current.

[4] The difference between 7.6×10^5 and 5.4×10^5 corresponds to a difference of about 18 percent in the wave-length, which is permissible in view of the approximate nature of our computations. Later we shall attempt to explain the causes of the difference between the theoretical and observed wave-lengths.

[5] E. Gill and T. Morrell, Phil. Mag. **44**, 161–178 (1922).

[6] For further details on different types of Barkhausen oscillations see H. E. Hollmann, Proc. I.R.E. **17**, 229–251 (1929); Zeits. f. Hochfr. **33**, 66–74, 101–105 (1929); K. Kohl, Erg. d. ex. Wiss. **9**, ch. II, §4 (1930).

[7] If one disregards the small variations in the wave-length produced by changes in the heating current.

TABLE III. *R5 (No. 2+No. 4). I_h=1.36A.*

L (cm)	E_g (volt)	I_a (mA)	λ (cm)	$\lambda^2 E_g$	L (cm)	E_g (volt)	I_a (mA)	λ (cm)	$\lambda^2 E_g$
				normal waves					
45	150	1.3	57.6	4.98×10^5	50	130	0.2	61.4	4.90×10^5
50	"	3.9	59.7	5.34 "	"	150	3.9	59.7	5.34 "
55	"	3.7	62.0	5.77 "	"	165	3.5	59.2	5.77 "
60	"	2.5	64.0	6.14 "	"	185	2.3	58.8	6.39 "
65	"	0.3	66.8	6.69 "	"	205	0.6	57.6	6.80 "
				dwarf waves					
12	100	0.2	42.8	1.83×10^5	17	90	0.1	45.6	1.87×10^5
17	"	1.6	44.9	2.01 "	"	100	1.6	44.9	2.01 "
19	"	0.3	45.8	2.10 "	"	110	0.2	44.4	2.17 "

The data of Table III clearly show that in either case the product $\lambda^2 E_g$ does not remain constant. Both for normal waves and for dwarf waves $\lambda^2 E_g$ varies in the same manner. The results this show unmistakably that both the normal waves and dwarf waves are GM-oscillations.

It must be mentioned, that for points which do not correspond to maximum values of the plate current the values of the products $\lambda^2 E_g$ differ from the average values found previously, namely, 5.44×10^5 for normal waves and 1.99×10^5 for dwarf waves. The difference, however, is not so great as when we pass from normal to dwarf waves. This shows that there is a fundamental difference between normal and dwarf waves. Before considering the causes of this difference, we shall discuss the results obtained with other tubes.

§5.

With the data of Table II and one of the working diagrams, it is easily found that dwarf waves are observed only in a very small region of oscillations, denoted with an arrow on Fig. 2. All the other regions of oscillations give normal waves. It is also seen from Fig. 2 and from the data of Table I that the region where dwarf waves are observed is of secondary importance as far as the place it occupies on the working diagrams and the energy of its waves are concerned.

In taking separate (I_a, E_g)-and (I_a, L)-characteristics one could easily fail to observe the dwarf waves. A complete picture of the work of a generator can be obtained only by constructing and thoroughly investigating working diagrams. The completeness of this method constitutes a very great advantage over all the other methods. The working diagrams have an additional advantage of being clear and convenient to use. The labor spent in constructing these diagrams is amply paid for.

It is of interest to know why the dwarf waves are observed in one small region only and not in several analogous regions as is the case for normal waves. Let us consider the diagram, Fig. 1, and let us take some point in the region of dwarf waves, point M^{I}, for example, which lies at $L=11$ cm and $E_g=110$ v. As seen from Table II, the waves generated at this point have a length $\lambda\sim42$ cm. According to what we said before (see §3), we should expect dwarf waves in points M^{II}, M^{III}, M^{IV}, lying at the same value of E_g and at

$L=11+\lambda/2$, $L=11+\lambda$, $L=11+3/2\lambda$. It is seen from the diagram that all these points lie in the regions of normal waves which are longer and therefore more easily generated. If we take some other point in the region of dwarf waves, point N^{I}, for example, which lies at $L=17$ cm and $E_g=100$ v and for which $\lambda\sim45$ cm, then similarly we might expect dwarf waves in points N^{II} and N^{III}. These points are also situated in the regions of normal waves. Hence it follows that if it were possible to inhibit the generation of normal waves in points M^{II}, M^{III}, M^{IV}, or N^{II}, N^{III} we could obtain dwarf waves at these points. This is easily verified experimentally.

So far, our plate and grid circuits were of equal length $L=L_a=L_g$ (see I, Fig. 1), i.e., they were in resonance. We shall now destroy this resonance. Let us take $L_a=32$ cm and $E_g=110$ v, corresponding to point M^{II}, and let us

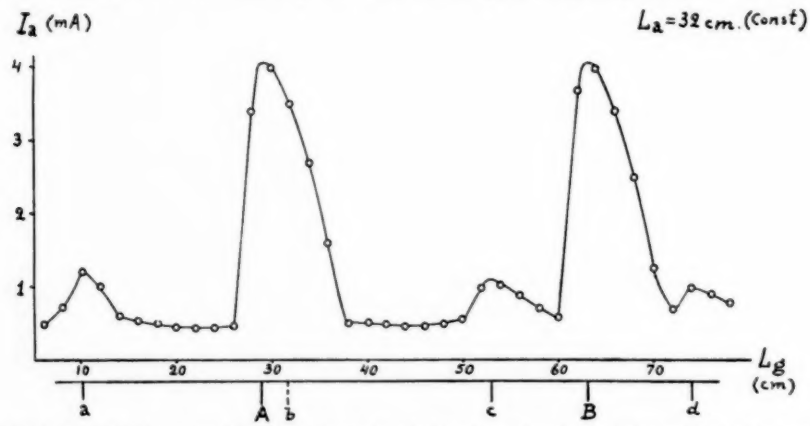


Fig. 4. (I_a, L)-characteristics, the lengths of the plate circuit and the grid circuit not being equal. The humps A and B correspond to normal waves, humps a, c and d, to dwarf waves.

vary the length L_g of the grid circuit, making simultaneous measurements of the plate current. The results of these observations are presented in the curve of Fig. 4. It is easily seen that the two main humps of this curve—A and B correspond to the normal waves. This is because at $L_g=32$ cm we should have normal waves[8] (at this value of L_g we have $L_a=L_g$) and also because the distance between the peaks of these two humps is approximately 33 cm, which is very nearly half the wave-length of the normal waves, which we should have at $E_g=110$ v (see Table II). It is easily shown that the three smaller humps, a, c, d, correspond to dwarf waves, since measurements show that at this value of L_g the average wave-length $\lambda=44.0$ cm. This is very near to the wave-length of dwarf waves, found previously at $E_g=110$ v. The distance between the humps c and d is 22 cm, i.e., is equal to a half wave-length. The distance between humps a and c is approximately 43 cm. It is easily guessed that between humps a and c there should be another hump b, which is not seen merely because it is concealed under the hump A.

[8] The peak of the hump A lies at $L_g\sim29$ cm, that is, at $L_g\neq L_a$. This is due to the fact that the point $L_a=L_g=32$ cm and $E_g=110$ v does not correspond to the maximum of the plate current.

Hence, it becomes clear that it is possible to inhibit the generation of normal waves if we put out of tune the plate and grid circuits. If we take $L_a = 32$ cm and $L_g = 53$ cm, or $L_a = 32$ cm and $L_g = 74$ cm and keep $E_g = 110$ v, we shall obtain dwarf waves, although, as we have seen, these dwarf waves are not observable either for $L_a = L_g = 32$ cm, or for $L_a = L_g = 53$ cm, or for $I_a = L_g = 74$ cm.

§6.

The above described method of *putting the circuits out of tune* is very useful when several kinds of waves are present at the same time and it is necessary to separate them. For instance, if at a certain circuit length $L_0 = L_a = L_g$

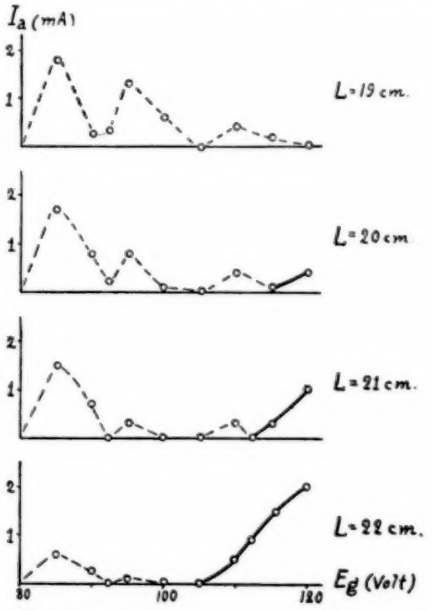


Fig. 5. (I_a, E_g)-characteristics showing the regions of dwarf waves at different circuit lengths;—normal waves; ---dwarf waves.

several kinds of oscillations are simultaneously generated with wave-lengths λ_1, λ_2, etc., we can evidently separate the first kind of waves by putting the circuits out of tune, making, for example, $L_a = L_0$ and $L_g = L_0 + n\lambda_1/2$. The second kind of waves can be separated at $L_a = L_0$ and $L_g = L_0 + n\lambda_2/2$ etc. Naturally, we can as well change the length of the plate circuit and keep the length of the grid circuit constant.

§7.

The regions of dwarf waves are not always of as simple an appearance as those seen on the preceding diagrams. Very frequently these regions consist of several parallel ridges merging gradually with a neighboring region of normal waves.

To discuss this case we shall consider several (I_a, E_g)-characteristics obtained with two other tubes of the same type R5. They are shown on Fig. 5.

Within the limits $L=19-22$ cm, $E_g=80-120$ v, i.e., where previously we had a region of dwarf waves, we see several regions of oscillations. As L increases these regions are gradually absorbed by the region of normal waves. The length of waves generated in different regions was determined at points corresponding to a maximum value of the plate crurent. Table IV gives the results of these observations for $L=19$ cm. It is seen from the table that for all the three regions $\lambda^2 E_g$ has approximately the same value, closely approaching that which was formerly obtained for dwarf waves. Measurements made at $L=20$, 21, and 22 cm gave for $\lambda^2 E_g$ values close to 2×10^5. Thus, there is no doubt that all the three regions belong to the same dwarf waves.

TABLE IV. *R5 (No. 3+No. 4). $I_A=1.36A$.*

E_g (volt)	L (cm)	I_a (mA)	λ (cm)	$\lambda^2 E_g$
85	19	1.8	48.3	1.98×10^5
95	"	1.3	46.2	2.03 "
110	"	0.4	42.2	1.96 "

The question as to why the region of dwarf waves is in the shape of a single ridge in some cases and in the shape of several ridges in other cases depending on the tube used, can be answered only after discussing the nature of oscillating circuits where dwarf waves originate. We shall not consider this problem now.

§8.

On the diagrams, shown above, all regions of oscillations appear very uniform and the lines of maxima (see Fig. 3) have neither gaps, nor sharp bends. This is not always the case. Sometimes the lines of maxima are sharply bent (see dotted lines on Fig. 3). In such a case typical depressions appear in the corresponding places of the 3-dimensional diagram. This is an indication that the oscillations generated at the corresponding values of E_g and L have a wave-length closely approaching one of the natural periods of the chokes and that the chokes no longer stop them. The energy of the oscillations is evidently dissipated in the connecting leads of the generator. The irregularities disappear when the choke is replaced by another one, having a different natural period. Thus, a regularly shaped diagram and the absence of bends in the lines of maxima is an indication that the generator is working correctly.

§9. COMPLEX WORKING DIAGRAMS. DWARF WAVES OF HIGHER ORDERS.

We have seen before that in the case of simultaneous operation of two tubes the secondary (with respect to the energy) regions of oscillations are concealed by the neighboring regions of more powerful oscillations (I,§8). As the preceding results show, the dwarf waves are generated exactly in these secondary regions. Therefore, in all the subsequent experiments only one tube was used for the production of waves. The other tube, "ballast" tube, was used as a condenser and for each tube investigated a special ballast tube was selected, as described earlier (I,§9).

For some of the tubes investigated diagrams were obtained that were considerably more complicated than those shown on Figs. 1–3. There appear a larger number of regions of oscillations. Also, for relatively small overheating there appear regions of oscillations at considerably higher grid potentials—potentials at which tubes heated normally do not generate any oscillations, as a rule.

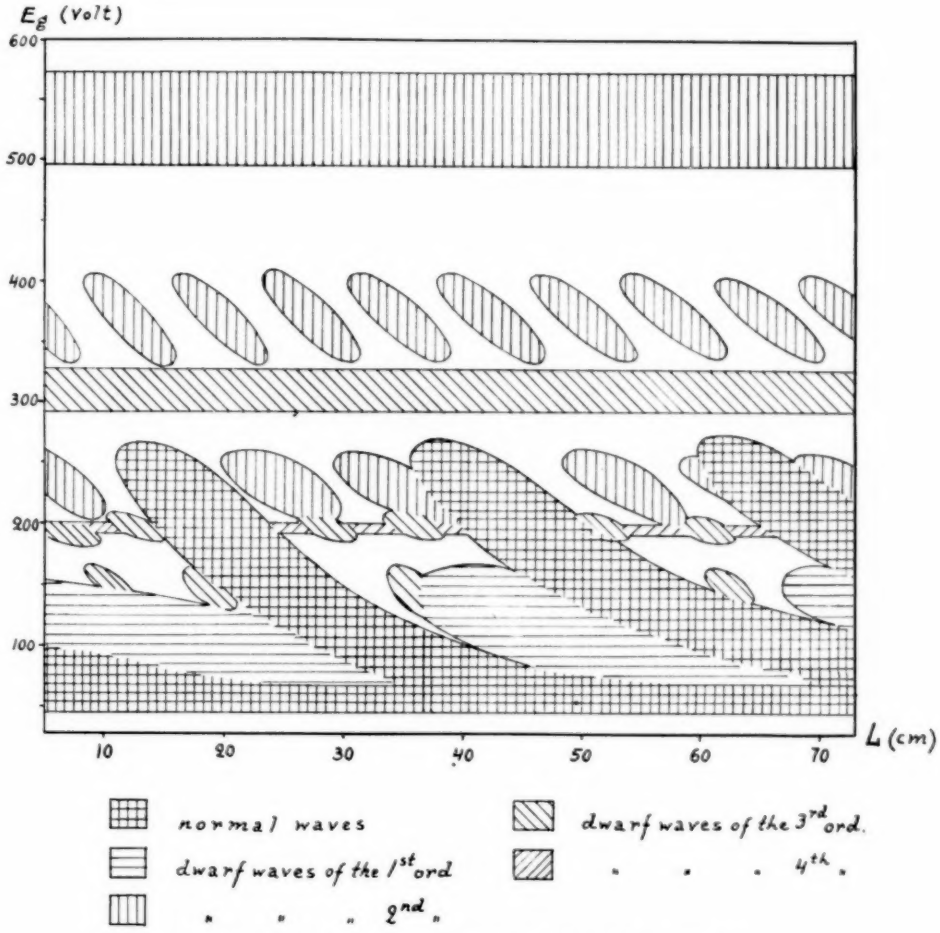


Fig. 6. Working diagram for vacuum tubes of the type *R*5, generating dwarf waves of higher order.

Fig. 6 shows a working diagram particularly rich in regions of oscillations. It was obtained with tube No. 20, type R5 and ballast tube No. 18 of the same type. The heating current was a little above normal so that the emission current reached as a maximum about 19 mA (normal $I_e = 5-7$ mA).

A comparison of this diagram with the diagram Fig. 1 shows the abundance of regions of oscillations. Still there is a similarity between the two diagrams. In the lower part of the new diagram, between $E_g \sim 50$ v and $E_g \sim 260$ v we observe the same principal regions of oscillations as before and

which, as we know, belong to normal waves. These regions are considerably narrower than before, which is as we should expect, as in this case only one tube was working and not two as before. Between the regions of normal weves we see a great number of smaller regions. Although the regions of normal waves are narrower and their energy is smaller, as the second tube is no longer working, still these regions reach to higher values of E_g than before. This is due both to the greater heating current (I, §2) and to the greater degree of

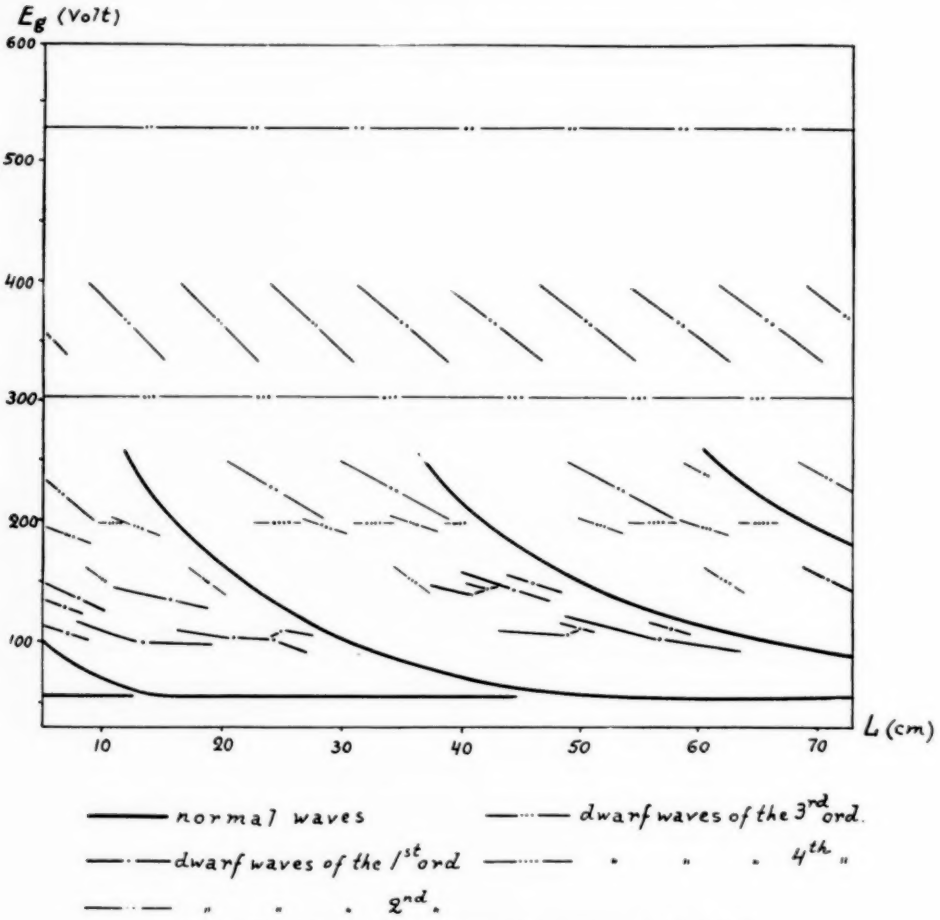


Fig. 7. Simplified working diagram for vacuum tubes of the type *R*5 generating dwarf waves of higher orders.

correspondence between the internal capacities of the two tubes, as it is found experimentally that the smaller the difference between the internal capacities of the working and the ballast tubes, the greater will be the length of the regions of normal oscillations. The lower parts of the regions of oscillations converge into a common asymptote. As we shall see later, this asymptote corresponds to the upper bend of the usual static characteristic of the tube.

We can also identify on Fig. 6 the region of dwarf waves. It lies in the left

lower corner of the diagram between $E_g \sim 100$ v and $E_g \sim 150$ v. It is considerably larger than formerly. Two other similar regions are seen within the same limits of E_g, the one between $L = 40 - 60$ cm and the other between $L = 65 - 72$ cm. These two regions must also belong to dwarf waves, as previously we expected to find them near $L = 50$ cm and $L = 70$ cm (see points M^{III}, N^{III} and M^{IV} Fig. 1).

The general properties of the regions of oscillations can easily be determined by means of the simplified diagram, shown on Fig. 7. Comparing it with the diagram of Fig. 6 we see that with the exception of the three regions of the dwarf waves, spoken of above, all the remaining regions have the same simple structure as those on the 3-dimensional diagram of Fig. 2. Each region has one ridge and a corresponding one line of maxima on the simplified diagram Fig. 7. Only the regions of dwarf waves mentioned above differ in appearance on account of the presence of several ridges going approximately in the same direction. The complex structure of the regions of dwarf waves has already been spoken of; now we see that this structure is characteristic just for these three regions.[9]

An inspection of the simplified diagram shows that the lines of maxima have different slopes. Among other things, this indicates that a change in the lengths of plate and grid circuits produces different effects on the oscillations of different regions. Of particular interest are the three regions of oscillations whose lines of maxima lie at $E_g \sim 200$ v, $E_g \sim 305$ v and $E_g \sim 530$ v. Unlike all the others, these regions go parallel to the axis of abscissas from one end of the diagram to the other. The width of the regions, and hence the amplitude of the corresponding waves, increases as E_g increases.

The (I_a, E_g)-characteristic of tube No. 20 with tube No. 18 as a ballast and at the circuits length $L = 50$ cm, was obtained previously (see I, Fig. 7). It gives an idea of the amplitude of the oscillations in different regions up to $E_g \sim 300$ v. For other regions, lying at higher values of E_g we can use the values of the maximum plate current given in Table V.

The distance between the regions of oscillations on Figs. 6 and 7 indicates that at about $E_g \sim 200$ v and $E_g \sim 360$ v tube No. 20 generates waves considerably shorter than those we had previously. This is most clearly seen in that part of the diagram near $E_g \sim 360$ v where the regions of oscillations are disposed with particular regularity.

Table V gives the results of measurements of wave-lengths in different regions of oscillations. As previously, these measurements were made at maximum values of the plate current.

It is seen from the table that there are five different kinds of regions of oscillations. For one of them the product $\lambda^2 E_g$ has an average value 5.78×10^5. The corresponding waves can be called normal. The second group gives for $\lambda^2 E_g$ an average of 1.97×10^5, which is very near to the value previously obtained for dwarf waves. The corresponding waves can be called *dwarf waves of the 1st order*. Waves of the remaining regions, for which the product $\lambda^2 E_g$

[9] We shall see later that they are regions of the so-called dwarf waves of the 1st order.

TABLE V. *R5 (No. 20). I_h=0.70 A.*

E_g (volt)	L (cm)	I_a (mA)	λ (cm)	λ_b (cm)	$\frac{\lambda_b}{\lambda}$	$\lambda^2 E_g$	E_g (volt)	L (cm)	I_a (mA)	λ (cm)	λ_b (cm)	$\frac{\lambda_b}{\lambda}$	$\lambda^2 E_g$
105	33	2.2	74.0	85.1	1.15	5.75×10^5	247	69	0.15	18.4	55.5	3.02	0.836×10^5
121	29	2.3	69.6	79.3	1.14	5.86	327	13	0.80	16.1	48.2	2.99	0.848
142	26	2.3	64.0	73.2	1.14	5.82	334	21	0.50	16.0	47.7	2.98	0.855
150	53	2.5	62.2	71.2	1.14	5.80	528	22	0.62	12.7	37.9	2.98	0.852
158	19	2.8	60.2	69.4	1.15	5.73	530	15	0.65	12.7	37.9	2.98	0.855
182	18	2.6	56.0	64.6	1.15	5.71	532	5	0.65	12.65	37.8	2.99	0.851
187	17	2.8	55.4	63.8	1.15	5.74							
190	71	2.6	55.0	63.2	1.15	5.75						$(\lambda^2 E_g)$	aver. $=0.846\times10^5$
196	16	2.5	54.0	62.3	1.15	5.72							
200	69	2.4	54.0	61.6	1.14	5.83							
200	15	2.5	54.2	61.6	1.14	5.88	140	6	0.50	18.8	73.7	3.92	0.495×10^5
							150	19	0.30	18.0	71.2	3.96	0.486
					$(\lambda^2 E_g)$	aver. $=5.78\times10^5$	160	9	0.30	17.35	68.9	3.97	0.481
							160	61	0.30	17.5	68.9	3.94	0.490
104	15	0.84	43.7	85.5	1.96	1.99×10^5	185	65	0.03	16.2	64.1	3.96	0.485
107	21	0.90	42.7	84.3	1.97	1.95	190	30	0.04	15.7	63.2	4.02	0.468
109	27	0.65	42.4	83.5	1.97	1.96	190	6	0.09	15.8	63.2	4.00	0.474
122	21	0.70	40.4	78.9	1.95	1.99	195	52	0.08	15.4	62.4	4.05	0.462
123	15	0.70	39.8	78.6	1.97	1.95	197	36	0.12	15.4	62.1	4.03	0.467
137	9	0.80	38.1	74.5	1.96	1.99	198	12	0.05	15.5	62.0	4.00	0.475
142	15	0.90	37.1	73.2	1.97	1.95	200	59	0.24	15.6	61.6	3.95	0.487
143	42	0.75	37.4	72.9	1.95	2.00	204	28	0.20	15.4	61.0	3.96	0.484
150	6	0.65	36.2	71.2	1.97	1.97	205	34	0.06	15.3	60.9	3.99	0.480
153	9	0.60	35.6	70.5	1.98	1.94	303	8	0.48	12.8	50.1	3.91	0.496
							303	21	0.45	12.65	50.1	3.96	0.485
					$(\lambda^2 E_g)$	aver. $=1.97\times10^5$	305	15	0.46	12.55	49.9	3.98	0.480
200	39	0.10	20.7	61.6	2.98	0.857×10^5						$(\lambda^2 E_g)$	aver. $=0.481\times10^5$
212	8	0.60	19.8	59.9	3.02	0.831							
237	72	0.53	18.8	56.6	3.01	0.838	200	33,55	0.20	12.4	61.8	4.97	0.308×10^5
243	70	0.28	18.6	55.9	3.01	0.841	330	13	0.04	9.4	48.0	5.10	0.292
												$(\lambda^2 E_g)$	aver. $=0.300\times10^5$

has a still smaller value, can be called *dwarf waves of the 2nd, 3rd and 4th orders*, depending on their length and the magnitude of the product $\lambda^2 E_g$. Dwarf waves of the 4th order are then the shortest. As seen from Table V some of them have a wave-length $\lambda = 9.4$ cm. These are the shortest waves we observed with tubes of the type *R*5.[10] No independent region of oscillations could be found for these waves and therefore it is plotted neither on the working diagram Fig. 6, nor on the simplified diagram Fig. 7. As we shall see later, these waves were always found on the boundaries of the regions of waves of other orders. These waves were first discovered with tube No. 20. They were found on the boundary of the dwarfs of the 2nd and the 3rd orders when observations were being made on the dependence of the wave-length of the dwarfs of the 3rd order on the grid voltage (see below). Subsequently these waves were observed with other tubes.

It is seen that the average values of $\lambda^2 E_g$ for normal waves and for dwarf waves of various orders differ widely. At the same time, for waves of the same order the individual values of the product $\lambda^2 E_g$ are very close to the average value. Thus, the determination of the order of a wave presents no difficulties.

Comparing the data of Table V with those of Table II it is seen that in the earlier experiments there is a greater difference between individual values of the product $\lambda^2 E_g$ for normal waves; the same is also true for the dwarf waves

[10] A. Wainberg, Journ. appl. Phys. (russ) **7**, 97–104 (1930) in an investigation made at our suggestion, was separating short waves by screening the tube by means of metallic shields. Using a tube of the type *R*5 he succeeded in obtaining waves 8 cm long at $E_g \sim 320$ v., i.e., where we found waves $\lambda = 9.4$ cm.

of the 1st order. This is due not so much to the lesser accuracy of the early measurements as to the fact that the early measurements were made at different times, frequently after considerable intervals of time and the tube was not always sufficiently preheated before the beginning of measurements (cf. I, §4).

Together with the observed wave-lengths λ Table V gives the wave-lengths λ_b which were calculated by means of Barkhausen's formula (2). It is seen that the length of the normal waves is close to the theoretical value.[11] The value of the ratio λ_b/λ shows that the dwarf waves of the 1st order are approximately twice as short as those given by formula (2); dwarf waves of the 2nd order are approximately three times shorter than the computed value; dwarfs of the 3rd order four times; and finally dwarfs of the 4th order are approximately five times shorter than those given by formula (2). Hence it follows that the wave-lengths of the dwarf waves are given by the following simple relationships:

$$\lambda_1 = \frac{\lambda_0}{2} \text{ (dwarf waves of the 1st order)}$$

$$\lambda_2 = \frac{\lambda_0}{3} \text{ (" " " " 2nd ")} \qquad (3)$$

$$\lambda_3 = \frac{\lambda_0}{4} \text{ (" " " " 3rd ") etc.,}$$

where λ_0 is the length of the normal waves, corresponding to the grid potential at which dwarf waves are observed. The approximate value of λ_0 can be computed from Eq. (2).

In the particular case it is evident that at the same grid voltage E_g we should observe a series of waves whose lengths correspond to series (3). Whether *all* these waves are actually generated or only some of them depends on the absolute value of λ_0 and on the presence of oscillating circuits, one of the natural periods of which is equal or near to the members of series (3). Without going into a detailed discussion we shall give two examples from the data of Table V. At $E_g = 150$ v the tube generates the following waves, $\lambda_0 \sim 62.2$ cm (normal waves), $\lambda_1 = 36.2$ ($\sim\lambda_0/2$) and $\lambda_3 = 18.0$ ($\sim\lambda_0/4$). At $E_g = 200$ v we have: $\lambda_0 = 54.1$ (normal waves), $\lambda_2 = 20.7$ ($\sim\lambda_0/3$), $\lambda_3 = 15.6$ ($\sim\lambda_0/4$) and finally $\lambda_4 = 12.5$ ($\sim\lambda_0/5$). It is seen that the observed waves do not exactly correspond to the series (3). We shall see later the causes of this discrepancy. In all cases without exception, the observed wave-lengths exceed the values expected from series (3) by approximately 15 percent. Hence it follows that the discrepancies between the calculated and the observed wave-lengths are not accidental but are brought about by some as yet unknown cause.

Formula (2) was derived on the assumption that the frequency of the

[11] The observed waves are approximately 15 percent shorter than those given by formula (2).

observed oscillations is equal to the frequency of the electronic oscillations within the tube. The existence of waves 2, 3, 4, etc., times shorter than those computed on the basis of this formula leads to the following very important conclusion: *a vacuum tube can generate oscillations whose frequency exceeds many times the frequency of electronic oscillations.*

§10.

We investigated the oscillations generated by tubes Nos. 2 and 4 and found that these oscillations can be considered as *GM*-oscillations, since the wave-lengths of normal and dwarf waves[12] depend on the grid voltage and on the dimensions of the grid and plate circuits. Comparing the working diagrams Figs. 6 and 7 with the diagrams of Figs. 1 and 3 it is seen that nearly all the regions of oscillations are of the same type. This indicates that the oscillations of all these regions depend in the same manner on the grid voltage and on the dimensions of the oscillating circuits and hence, that they must be *GM*-oscillations.

There are, however, on the diagrams, Figs. 6 and 7, three regions different from the rest. These are regions going parallel to the axis of abscissas and having lines of maxima approximately at $E_g = 200$, 305 and 530 volts. From the data of Table V it is seen that oscillations generated within these regions have approximately the same wave-length, irrespective of the dimensions of the oscillating circuits used. The latter circumstance is very important, for we know that one of the characteristic features of the *BK*-oscillations is the fact that their wave-length is independent of the length of the oscillating circuits.

In order to decide whether or not the oscillations of these three regions are *BK*-oscillations, we must determine how their wave-length depends on the grid potential. If they are BK-oscillations, then a change in the grid potential must produce a change in the wave-length according to the Eq. (1). Table VI

TABLE VI. *R5* (*No. 20*). $I_h = 0.70A$.

E_g (volt)	L (cm)	I_a (mA)	λ (cm)	$\lambda^2 E_g$	E_g (volt)	L (cm)	I_a (mA)	λ (cm)	$\lambda^2 E_g$
285	15	0.06	12.9	0.474×10^5	490	15	0.01	12.8	0.803×10^5
295	"	0.28	12.7	0.476 "	500	"	0.04	12.8	0.819 "
305	"	0.46	12.55	0.480 "	510	"	0.20	12.75	0.829 "
315	"	0.32	12.4	0.484 "	520	"	0.35	12.75	0.846 "
325	"	0.06	12.3	*0.492* "	530	"	0.65	12.7	0.855 "
330	"	0.02	9.4	*0.292* "	540	"	0.40	12.7	0.871 "
335	"	0.01	16.2	0.878 "	550	"	0.30	12.7	0.887 "
					560	"	0.20	12.65	0.896 "
					570	"	0.02	12.65	0.912 "

gives the results of measurements of wave-lengths at different grid potentials. Measurements were made for two regions only. No data are given for the third region because the narrowness of that region and its close proximity to the regions of other waves introduces an uncertainty in the measurements.

[12] They were dwarf waves of the 1st order.

Table VI shows that the wave-length changes with the grid potential. The nature of that change is quite analogous to that previously obtained for normal and dwarf waves of the 1st order. The product $\lambda^2 E_g$ changes when the grid potential is changed, which shows that these oscillations cannot be considered BK-oscillations. This conclusion can be extended to the oscillations of the third region, lying at $E_g \sim 200$ v as there is no reason to consider it different from the other two.

The results obtained permit us to assert that the observed *normal waves and all the dwarf waves of all orders belong to the same type of GM-oscillations.* It is of interest to know why the oscillations in these three regions are different from all the others and why their wave-length does not appreciably depend on the lengths of the plate and of the grid circuits. This can be readily understood if we assume that a circuit with a natural wave-length close to $\lambda = 12.5$ cm, in which these oscillations originate, lies somewhere within the tube[13] and that the plate and grid circuits are coupled with this circuit so that their oscillations are forced oscillations. In such a case a change in the

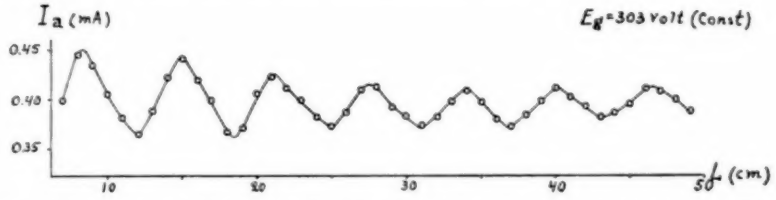


Fig. 8. (I_a, L)-characteristic showing the dependence of the amplitude of oscillations of dwarf waves of the 3rd order on the length of the grid and plate circuits.

wave-length of the oscillations in the primary circuit, which depends on the change of grid potential, will produce a corresponding change in the wave-length of the oscillations in the plate and grid circuits. As the oscillations in the primary circuit are not affected by the forced oscillations, a change in the length of the plate and grid circuits cannot produce a change in the wave-length of the oscillations in the primary circuit. This is what we observe.

At present it is difficult to say what is the cause of the dependence of the wave-length in the primary circuit on the grid potential and the problem requires a special investigation. It is clear, however, that this dependence cannot be entirely ascribed to a change in the time of passage of the electrons, since Eq. (1) is not satisfied when E_g changes.

If our assumption is correct, that there exists within the tube a circuit with a natural wave-length $\lambda = 12.5$ cm and that the plate and grid circuits are coupled with it, then a change in the length of these circuits must produce a change in the amplitude of the oscillations. The curves of Fig. 8 shows that this is verified by experiment. This curve represents the relation between the plate current and the lengths of the plate and grid circuits. The grid voltage was kept constant at a value corresponding to the line of maxima of the second (i.e., the middle one) region of oscillations. The curve, as we see, is

[13] As we shall see later, this assumption is supported by a number of other considerations.

the usual curve of standing waves.[13a] The wave-length found from this curve is very near to the value found by direct measurements.

§11.

During the above described measurements we succeeded in discovering waves $\lambda = 9.4$ cm, which are the shortest ones we found with tubes of the type *R*5. As shown in Table VI, with a change of grid voltage from $E_g = 325$ v to $E_g = 330$ v the wave-length decreased abruptly from $\lambda = 12.3$ cm to $\lambda = 9.4$ cm, with a simultaneous decrease in the plate current. Such an abrupt transition from one kind of waves to another is usually accompanied by transitions from one region of oscillations into a neighboring one. At $E_g = 330$ v we did not succeed in finding a special region of dwarf waves of the 4th order to which waves $\lambda = 9.4$ cm belong. When E_g was increased by $3-5$ v, the wave-length increased abruptly to $\lambda = 16.2$ cm which corresponds to the boundary of the region of dwarf waves of the second order which exists in this part of the diagram. The waves $\lambda = 9.4$ cm disappeared when the length of the plate and grid circuits was changed by $1\frac{1}{2} - 2$ cm. They reappeared at $L = 6-7$ cm and $L = 23-24$ cm, that is, again in those points where the region of dwarf waves of the 2nd order passes into the region of dwarfs of the 3rd order.

Subsequently, waves near to $\lambda = 9.5$ cm were obtained with other tubes. For example, tube No. 5 gave waves $\lambda = 9.4$ cm. These observations shall be presented later. Such waves were obtained at $E_g = 330$ v and at $E_g = 335$ v. In both cases we had $L = 13.5-15$ cm, i.e., in both cases the waves were found on the boundary of the dwarfs of the 2nd and 3rd orders. Again, in the very beginning of our investigation, tube No. 2 gave waves $\lambda = 9.6$ cm at $E_g = 90$ v and $L = 35-40$ cm.[14] The working diagrams show that this point corresponds, to the boundary of the region of normal waves.

All these observations show that waves approximately 9.5 cm long are characteristic for tubes of the type *R*5. They also indicate that the appearance of these waves is in some way connected with the unstable "regime" existing in the tubes at the boundaries of regions of oscillations. This should be mentioned because other types of tubes give sometimes very short waves at unstable operating "regimes."

§12.

The working diagrams show that tube No. 20 began generating oscillations at $E_g \sim 55$ v. The measurements given in Table V do not go below 105 v. The discrepancy is not accidental. At the heating current at which tube No. 20 was operated, the upper bend of the usual static characteristic of that tube was located between $E_g = 50-80$ v. Thus all the points of the diagram for which $E_g < 80$ v went beyond the region of the saturation current. Since the space charge must have a distorting influence in all points outside of the region of saturation current, all these points were excluded from our observations.

[13a] Cf. M. O. J. Strutt, Ann. d. Physik **4**, 26 (1930).

[14] In this case $\lambda^2 E_g = 0.083 \times 10^5$. It is evident that we are dealing with dwarf waves of a fairly high order (8^{th} or 9^{th}).

Curve (a) Fig. 9 gives an idea of the results obtainable in the region $E_g = 50 - 80$ v. The dotted line corresponds to the equation $\lambda^2 E_g = 5.78 \times 10^5$. It is seen that all points within the region $E_g = 105$–200 v (former observations) fit the curve very well, while within the region $E_g = 55 - 80$ v the points follow an entirely different curve. As a result the product $\lambda^2 E_g$ for these points differs considerably from the average value found previously, as is seen from curve (b) Fig. 9 and also from the data given in Table VII.

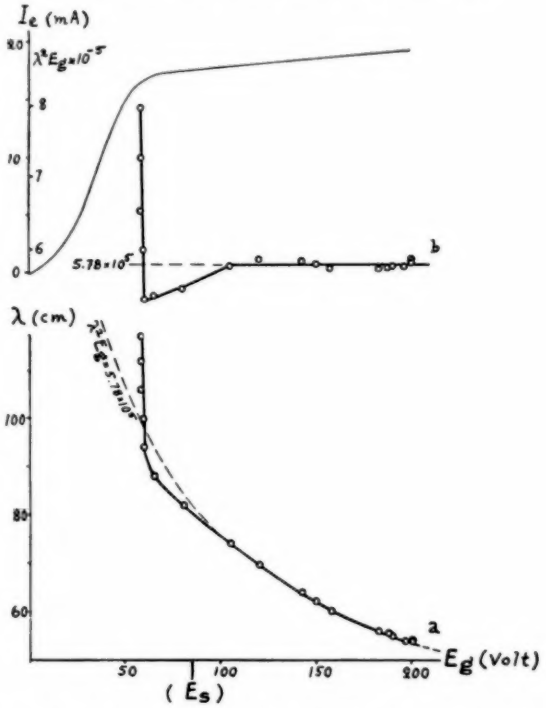


Fig. 9. Curves showing the dependance of the length of the normal waves, the value of the product $\lambda^2 E_g$ and the emission current on the grid potential. The dotted line (a) corresponds to Barkhausen's Eq. (1).

It is seen that at grid voltages $E_g < 100$ v, the wave-length is below that given by the Eq. $\lambda^2 E_g = 5.78 \times 10$ and the value of the product $\lambda^2 E_g$ falls off rapidly. This phenomenon is due to the fact that space charges appear near the filament. Analogous results were obtained by N. Kapzov and S. Gwosdower[15] who worked with vacuum tubes of a different type. Our observations shows that the product $\lambda^2 E_g$ has a minimum value at $E_g = 60$ v and $L = 50$ cm when the line of maxima comes close to its asymptote which lies at $E_g = 58$ v at the heating current used. In the region of this asymptote we have a very interesting phenomenon—at constant E_g the wave-length increases as the length of the circuits is increased (see Table VII) and the corresponding points of curves Fig. 9 lie above the dotted lines. As the grid voltage remains practically unchanged, the case is evidently quite analogous to the one we

[15] N. Kapzov u. S. Gwosdower, Zeits. f. Physik **45**, 114–134 (1927).

TABLE VII. *R5* (*No. 20*). $I_a = 0.70A$.

E_g (volt)	L (cm)	I_a (mA)	λ (cm)	$\lambda^2 E_g$
105	33	2.2	74	5.75×10^5
81	39	2.0	82	5.45 "
65	45	1.9	88	5.34 "
60	50	1.3	94	5.30 "
60	55	0.97	100	6.00 "
58	60	0.77	106	6.52 "
58	65	0.63	112	7.27 "
58	70	0.53	117	7.95 "

had previously, when the wave-length depended on the lengths of the grid and plate circuits—see Table III. In the present case, however, an increase in the circuit length produces a considerably greater increase in the wave-length than before. This is apparently due to the fact that at $E_g \sim E_s$ the oscillations are considerably less stable than at $E_g > E_s$ and all the external factors exercise a greater influence.

§13.

Measurements made with other vacuum tubes of the same type *R*5 gave results very similar to those described above.

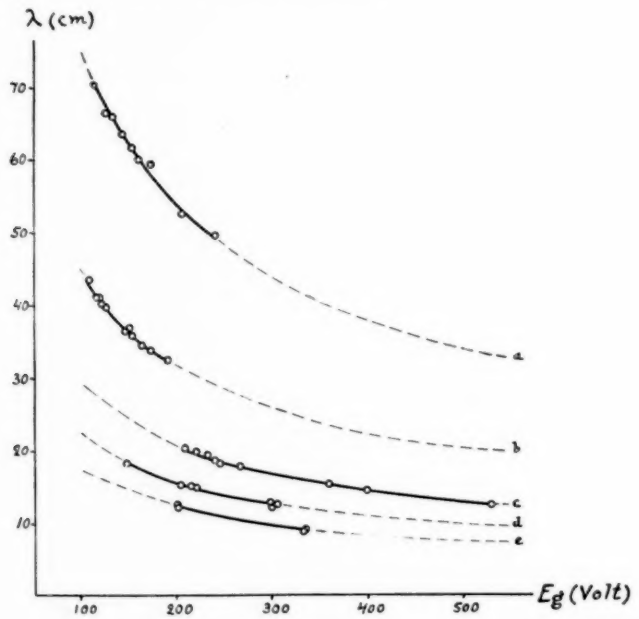


Fig. 10. The lengths of normal and dwarf waves at different grid voltages. The dotted lines correspond to the equation $\lambda^2 E_g = \text{Const}$. Curve (a) $-\lambda^2 E_g = 5.81 \times 10^5$ (normal waves); curve (b)—$\lambda^2 E_g = 2.01 \times 10^5$ (dwarf waves of the 1st ord.); curve (c) $-\lambda^2 E_g = 0.856 \times 10^5$ (dwarf waves of the 2nd ord.); curve (d)—$\lambda^2 E_g = 0.495 \times 10^5$ (dwarf waves of the 3rd ord.); curve (e)—$\lambda^2 E_g = 0.303 \times 10^5$ (dwarf waves of the 4th ord.).

We shall give as an example the results obtained with tube No. 5 (type *R*5). Tube No. 20 which has been investigated earlier, was used as a "ballast." All

measurements were made at a higher heating current than before, the emission current reaching as high as 41 mA. All the measurements were made within a very brief period of time, as at this heating current the disintegration of the filament proceeded rapidly.

We shall omit the tables of data and the working diagrams which are very similar to those of Figs. 6 and 7. Fig. 10 gives the results of measurements of wave-length in different points of the diagram, corresponding to the maximum values of the plate current. It is seen that the observations fit well the curves $\lambda^2 E_g = \text{const}$. The curves differ in the numerical value of the constant. The numerical values are very close to those previously obtained with tube No. 20. We are evidently dealing with five different kinds of waves: normal waves and dwarf waves of the 1st, 2nd, 3rd and 4th orders, corresponding to the numerical values of $\lambda^2 E_g$.

It could be easily shown that the waves generated by tube No. 5 almost coincide with those generated by tube No. 20. They lie now at slightly higher values of E_g than before. In other words, a comparison of the working diagrams for these two tubes would show that the regions of oscillations of tube No. 5 lie above those of tube No. 20. We shall see later that such a translation of regions is a result of the increase in the heating current.

We shall mention the following details. Due to the increase in heating current, the region of normal oscillations reaches now to higher values of E_g than before. On the curve of normal oscillations (Fig. 10) we see points up to $E_g = 242$ v, the boundaries of the region reaching almost up to $E_g = 300$ v. The region of dwarf waves of the second order, lying near $E_g \sim 350$ v, shows an appreciable increase in size, reaching up to $E_g = 400 - 410$ v. All the other regions remained practically unchanged. Observations with tubes Nos. 5 an 20 were made at a relatively high value of the heating current. We shall give also the results obtained with tube No. 7 (type *R*5), when the heating current was considerably lower, so that the emission current was about 12 mA.[16] Exactly similar results were obtained, the average value of $\lambda^2 E_g$ being: 5.85×10^5 for normal waves, 1.90×10^5 for dwarf waves of the 1st order, 0.800×10^5 for dwarf waves of the 2nd order, 0.466×10^5 for dwarf waves of the 3rd order, and, finally, 0.300×10^5 for dwarf waves of the 4th order. These values agree well with those found previously.

All our findings point to the following conclusion: as far as the generation of ultra-short waves is concerned, tubes of the same type differ only with respect to the number and the size of the regions of oscillations on the working diagrams. Moreover, if a region exists it will be located in nearly the same part on all the diagrams and the corresponding wave-lengths and values of $\lambda^2 E_g$ will be nearly equal for all the tubes. It should be borne in mind that observations should be made at a constant heating current and in those points of the diagram for which $E_g > E_s$ and which correspond to the maximum values of the amplitude of oscillation (maximum plate current).

[16] The results of our earlier observations, see Zeits. f. techn. Physik **10**, 542–548 (1929), Table III, are not given here, because they are less accurate and because tubes of the older series were used which had a slightly different diameter of the plate.

There is one exception to this rule. It occurs in those infrequent cases when two kinds of waves are generated at the same time and we have simultaneously dwarf waves and weak normal waves. This is sometimes observed in those regions of dwarf waves, which lie at $E_g < 300$ v, i.e., where normal waves can generally be observed. In such a case the wave-length and the value of $\lambda^2 E_g$ are higher than usual. An explanation of this fact can be given, after an investigation is made of the dependance of the length of the waves on the heating current and the amplitude of oscillation.

§14.

It is of importance to know, why some tubes have working diagrams rich in regions of dwarf waves (Figs. 6 and 7) and generate such waves easily, while other tubes of the same type have diagrams almost devoid of regions of dwarf waves and these waves are generated with difficulty.

In the case of normal waves it is known that their production is facilitated by a symmetry in the construction of the tube. This fact is supported by direct observations on the generation of these waves by tubes with destroyed symmetry[17] and also by the well-known fact that normal waves are nearly always generated by tubes with cylindrical electrodes, while tubes with plane electrodes produce them only in exceptional cases. The effect of symmetry is easily understood. If the symmetry of a tube with cylindrical electrodes is destroyed, i.e., the filament does not coincide with the axis of the grid and of the plate, then the time required for the electrons coming off the filament to pass to the plate will be different for different electrons, depending on the direction of their motion. If the difference in the time of passage is sufficiently large in comparison with the period of the generated oscillations, the oscillations will be broken up. Hence it follows that the condition of symmetry is essentially *a condition of the equality of the times of passage of the electrons* in different directions within the tube. In the case of a tube with cylindrical electrodes of the usual construction, i.e., with a filament inside of the grid, the two conditions are equivalent. In the case of tubes of some other design it is necessary and sufficient to satisfy the second condition only.

In a recent paper H. E. Hollmann[18] has indicated the possibility of obtaining oscillations using asymmetric tubes, with the anode placed within the cylindrical grid and the filament (one) located outside of the grid. This is not in contradiction with our statement. In these tubes the majority of electrons will follow approximately the direction of the straight line connecting the filament and the plate. Thus the condition of the equality of times of passage will be satisfied for the majority of electrons, in spite of the asymmetry of the construction. Our statement is further supported by the well-known fact that oscillations can be obtained with cylindrical tubes having spiral shaped filaments.

The condition of the equality of the times of passage must also have an influence on the generation of dwarf waves. Also, the higher the order of the

[17] See, for example, G. Breit, J. Opt. Soc. Am. **9**, 907–922 (1924).

[18] H. E. Hollmann, Phys. Zeits. **31**, 56–63 (1930).

dwarf waves, the more they are affected by an inequality in the times of passage, because the greater will be the relative difference between the period of oscillation and the times of passage of the electrons in different directions. Following our suggestion A. Wainberg[19] has systematically investigated the effect of the position of the filament in tubes of the type *R*5 on the generation of dwarf waves. These investigations[20] show that shifting the filament by 0.5 mm from the axis of the plate produces a decrease in the amplitude of the normal waves from 10 percent to 30 percent depending on their length and completely prevents the generation of dwarf waves of the 1st and 2nd orders[21] which normally are generated by the tube.

These results explain why working diagrams obtained with tubes of the same type can differ so widely with respect to the number of regions of dwarf waves.

For tubes of the type *R*5 we can calculate, using formula (2), that displacing the filament by 0.5 mm relative to the plate will cause a maximum difference of 20 percent in the times of passage of two electrons moving in opposite directions along the same diameter. This difference amounts to 40 percent of the period of the dwarf waves of the 1st order and to 80 percent of the period of the dwarf waves of the 2nd order. This explains why a small displacement of the filament has such a great effect in preventing the generation of dwarf waves.

In the commercially available tubes the electrodes are frequently displaced from their normal position by more than 0.5 mm. Therefore, very few of these tubes are capable of generating dwarf waves and especially dwarf waves of higher orders.

§15. On the Nature of Dwarf Waves

Series (3) is very similar to the series of overtones of harmonic oscillators. The resemblance is purely superficial, however, and dwarf waves cannot be considered as overtones of the normal waves.

The main difference between dwarf waves and overtones, as they are usually understood, consists in the fact that dwarf waves are completely independent of the normal waves. This is shown in Figs. 6 and 7. Dwarf waves are observed, in fact, even at such grid potentials E_g at which no normal waves can be produced at all. For example, at grid voltages of 300 v and higher, at which we have no traces of normal waves we can easily obtain dwarf waves of the 2nd and 3rd orders.

This does not mean, of course, that the wave-length of the dwarfs may not coincide with that of one of the overtones of the grid and plate circuits. All the points of lines of maxima will always correspond to one of the natural

[19] A. Wainberg, Journ. appl. Phys. (russ) **8**, 97–104 (1930).

[20] In tubes of the type *R*5 the plate is so arranged that it can easily be shifted from its normal position. Thus, the same tube could be used in all measurements. This guaranteed constancy of vacuum, filament, etc.

[21] A. Wainberg, reference 19, see Figs. 3 and 4, region IV (dwarf waves of the 1st order) and region V (dwarf waves of the 2nd order).

periods of these circuits (fundamental or overtone) and this coincidence may occur. But in this case the oscillations will not have the corresponding fundamental period, i.e., they will be forced.

§16.

According to Barkhausen's theory, the period of the oscillations in the circuits coupled with the tube must be equal to the period of the oscillations of the electrons about the grid. In other words, it must be equal to the time required for the electrons to pass from the filament to the plate and back. The derivation of Eq. (2) is based on this assumption. Hence it follows that the lengths of the normal and of the dwarf waves can be represented by the following relationships, which are analogous to the relationships (3):

$$\begin{aligned}
\lambda_0 &= c_0\tau \quad \text{(normal waves)} \\
\lambda_1 &= c_0\frac{\tau}{2} \text{ (dwarf waves of the 1st order)} \\
\lambda_2 &= c_0\frac{\tau}{3} \text{ (“ “ “ “ 2nd “)} \qquad (4) \\
\lambda_3 &= c_0\frac{\tau}{4} \text{ (“ “ “ “ 3rd “) etc.,}
\end{aligned}$$

where c_0 is the velocity of light and τ the time required for an electron to pass from the filament to the plate and back, corresponding to the value of the grid potential at which waves are observed.

If the periods of the oscillations generated be denoted by T_0 (normal waves), T_1 (dwarf waves of the 1st order), T_2 (dwarf waves of the second order) etc., series (4) gives

$$\tau = T_0 = 2T_1 = 3T_2 = 4T_3 = \cdots . \qquad (5)$$

This means that during τ, the time of electronic passage, the oscillating circuit can perform not one oscillation only, but two, three, four and so on.

This answers the question regarding the nature of dwarf waves. Since dwarf waves are GM-oscillations and since oscillations of this type orginate in the circuits coupled with the tube, it is clear that *dwarf waves are oscillations of the circuits within the tube or coupled with the tube and which are excited in such a maner that during the time of passage of electrons from the filament to the plate and back, they perform two complete oscillations (dwarf waves of the 1st order) three complete oscillations (dwarf waves of the 2nd order) etc.*

The difference between the normal and dwarf waves lies in the difference in the excitation of the oscillating circuits. The question as to how the circuits are brought into oscillations corresponding to dwarf waves is evidently analogous to the question as to the mechanism of the excitation of normal waves. The solution of this question involves an investigation of the mechanism of the electronic oscillations in a vacuum tube. In a preliminary paper[22] we have indicated one of the possible ways.

[22] G. Potapenko, Zeits. f. techn. Physik **10**, 542–548 (1929).

There remain some small systematic discrepancies between the results of observations and series (3) and therefore series (4) and (5). An investigation of the relation between the wave-length and heating current and the amplitude of the oscillations will explain the cause of these discrepancies.

§17.

We shall now consider to what regions of the spectrum will belong the normal and dwarf waves generated by tubes of the type *R*5. Fig. 11, which is based on the data of Fig. 10 and of Table V, shows that the oscillations generated cover only certain regions of the spectrum. Between these regions there are considerable gaps where no oscillations are generated at all. The normal waves cover a region of the spectrum corresponding to a wave-length $\lambda \sim 50$ cm and longer. The dwarf waves of the 1st order cover a region between $\lambda \sim 34$ cm and $\lambda \sim 44$ cm. Dwarf waves of the 2nd, 3rd and 4th order lie within still narrower limits around $\lambda = 18.5$ cm, $\lambda = 15.5$ cm, $\lambda = 12.5$ cm and $\lambda = 9.5$ cm. It is of interest that waves close to $\lambda = 18.5$ cm and $\lambda = 15.5$ cm are generated

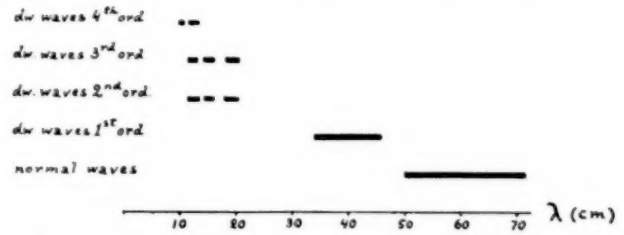


Fig. 11. Regions of the spectrum corresponding to the oscillations generated by vacuum tubes of the type *R*5.

both as dwarf waves of the 2nd order and as dwarf waves of the 3rd order. Waves close to $\lambda = 12.5$ cm are generated as dwarf waves of either 2nd, 3rd or 4th orders.

The fact that only certain waves whose length lies within definite limits can be generated as dwarf waves of the 2nd, 3rd and 4th orders shows that somewhere within the tube there exist oscillating circuits whose natural periods (fundamental or overtones) are close to the periods of the observed oscillations. These circuits exist somewhere within the bulb or the base of the tube, because the length of the corresponding waves (waves near to $\lambda = 12.5$ cm) is little or not at all dependent on the length of the grid and plate circuits. The question as to what these internal oscillating circuits really are is rather complicated and must be treated separately. In connection with this question we had constructed a special tube in which one of the internal circuits was variable. The design of this tube was described in a previous paper.[23]

We have seen that waves of the same length can be generated as dwarf waves of different orders. This is in complete agreement with our previous considerations. It shows that the same circuit can be made to oscillate in several different ways, so that during the passage of electrons from the fila-

[23] G. Potapenko, reference 22, see Figs. 7 and 8.

ment to the plate and back to the filament it will make three, four or five oscillations depending on whether we have dwarf waves of the 2nd, 3rd or 4th orders respectively.

From Eq. (1) it is easily seen that the values of the grid potentials at which waves of the same length can be observed must satisfy the following series

$$(E_g)_0, \frac{(E_g)_0}{4}, \frac{(E_g)_0}{9}, \frac{(E_g)_0}{16}, \ldots \tag{6}$$

where the first term corresponds to normal waves, the second to dwarf waves of the 1st order, etc. The correctness of the above series can be seen from the data of Table V.

Fig. 11 shows that waves corresponding to the oscillations of the internal circuits of the tube can be observed as dwarf waves only and only as dwarf waves of certain definite orders. For example, none of them appear as dwarf waves of the 1st order. As dwarf waves of the 4th order we have waves near to $\lambda = 9.5$ cm and $\lambda = 12.5$ cm. We do not find here any waves of greater length which, one would think, could be more easily obtained. An explanation of this can be obtained from a consideration of Table VIII, which gives the grid potentials at which we can expect the appearance of waves corresponding to internal circuits in the form of dwarf waves of different orders. These potentials are calculated using formula (1), the numerical values of the product $\lambda^2 E_g$ being taken from Table V. Voltages at which waves were actually observed are separated by a thick line from the other figures.

TABLE VIII.

Order of the waves	$\lambda^2 E_g$	$\lambda = 9.5$	$\lambda = 12.5$	$\lambda = 15.5$	$\lambda = 18.5$
Dwarf waves 1 ord.	$1.97 \cdot 10^5$	2200 v	1300 v	820 v	580 v
" " 2 "	0.846 "	940 "	540 "	450 "	250 "
" " 3 "	0.481 "	530 "	310 "	200 "	140 "
" " 4 "	0.300 "	330 "	190 "	120 "	90 "
" " 5 "	$(0.21 \cdot 10^5)$	230 "	130 "	90 "	60 "

From Table VIII it is seen that waves $\lambda = 15.5$ cm and $\lambda = 18.5$ cm could be generated as dwarf waves of the 4th order at grid potentials $E_g = 120$ v and $E_g = 90$ v. The working diagram Fig. 6 shows, however, that longer and more powerful normal waves and dwarf waves of the first order are generated at these voltages. Their presence precludes the appearance of the dwarf waves of the 4th order which are shorter and hence generated with greater difficulty. Waves $\lambda = 9.5$ cm could be obtained as dwarf waves of the 3rd order at a grid potential $E_g = 530$ volts. This is hindered, however, by the longer waves $\lambda = 12.5$ cm, which, as we have seen before, occupy the whole region of the diagram which lies between $E_g \sim 500 - 550$ v. We cannot expect to obtain waves $\lambda = 9.5$ cm as dwarf waves of the 2nd order, in view of the high potential required. For the same reason, evidently no dwarf waves of the first order $\lambda = 9.5 - 18.5$ cm were observed.

At $E_g = 190-200$ v it is theoretically possible to obtain dwarf waves of the 4th order $\lambda = 12.5$ cm and dwarf waves of the 3rd order $\lambda = 15.5$ cm. These waves were actually observed. In this case their appearance was not prevented by longer waves. This is explained by the fact that waves $\lambda = 15.5$ cm are generated only at certain definite lengths of the plate and grid circuits, while waves $\lambda = 12.5$ cm are independent of the length of the circuits. Thus at certain circuit lengths the waves $\lambda = 12.5$ can appear as they are not hindered by the longer waves. That is where they were observed.

From series (3) it is easy to calculate the probable value of $\lambda^2 E_g$ and the potentials at which dwarf waves of the 5th order could be generated. These calculations are also given in Table VIII. Comparing them with the working diagram Fig. 6 it would seem that waves $\lambda = 9.5$ and especially waves $\lambda = 12.5$ might be observed as dwarf waves of the 5th order. However so far all attempts to find them were unsuccessful. Apparently, this is due to the fact that a higher degree of symmetry in the arrangement of the electrodes is required for the generation of dwarf waves of higher orders.

The production of dwarf waves of higher orders is of interest because it enables us to obtain short waves at relatively very low voltages. The latter ensures a steadiness in the operation of the tube and renders possible the application of short waves for exact measurements. A few simple computations will show the advantages of dwarf waves as far as a decrease in grid voltage is concerned. To obtain normal waves $\lambda = 18.5$ cm a grid voltage $E_g \sim 1900$ v is required. To obtain normal waves $\lambda = 9.5$ cm, a grid voltage $E_g \sim 6400$ v is necessary. These figures are so large that the production of normal waves of this length is entirely out of the question. Besides, at such values of the grid voltage the emission current would have to be increased approximately 150 times (see I, §2). This is also quite impossible.

In conclusion the author wishes to express his gratitude to the Rockefeller Foundation for the grant of a Fellowship and to Professor R. A. Millikan for the facilities of the Norman Bridge Laboratory.

A DETERMINATION OF THE TOTAL ELECTRIC POLARIZATION AND THE ELECTRIC MOMENTS OF CERTAIN ORGANIC MOLECULES

By Louis M. Heil
Mendenhall Laboratory of Physics, Ohio State University

(Received September 14, 1931)

Abstract

An investigation has been made of the effect of temperature on the dielectric constants and densities of bromobenzene and iodobenzene in dilute solution with hexane and of hexane, and of m-p-o-xylenes as pure liquids.

The values of the electric moments and the validity of the Debye law: The values obtained for the electric moments are as follows; p-xylene, 0; m-xylene, 0.34×10^{-18}; o-xylene, 0.44×10^{-18}; bromobenzene, 1.35×10^{-18}; iodobenzene, 1.25×10^{-18}. In every case it was found that the Debye law, $P=P_0+a/T$, was valid.

Variation of the electric moment with variation of the atomic weight of a substituted atom; The results of this investigation give additional evidence of the rule that, as the atomic weight of a substituted atom is increased, the electric moment of the molecule decreases.

Change of induced polarization with the temperature. Evidence has been deduced which suggests that the induced polarization may be a function of the temperature and that this effect may neutralize or diminish the effect of a small dipole.

Introduction

THE total polarization of a molecule is generally assumed to consist of three parts; (1) the electronic polarization, which is ascribed to the oscillatory motion of the electrons in the atom under the action of the alternating field of the exciting radiation, (2) the atomic polarization, produced by the vibration of the atoms or ions under the action of the alternating field, (3) the dipole polarization, which is ascribed to an electrically unbalanced molecule which is free to oscillate as a whole.

According to the Debye theory the total polarization P is,

$$P = P_e + P_a + P_d \tag{1}$$

where P_e is the electronic polarization; P_a the atomic polarization; P_d the dipole polarization. $P_d=(4/9)\pi Nu^2/kT$ where N is Avogadro's number; u the electric moment of the molecule; k the Boltzmann constant per molecule. The value of the total polarization can be obtained, after a measurement of the dielectric constant and density of a substance at a certain temperature, from the Clausius-Mossotti relation,

$$P = \frac{e-1}{e+2}\times\frac{M}{d}. \tag{2}$$

This total polarization P may be measured at any temperature T by a method like that used by Smith,[1] Stranathan,[2] or Williams.[3] From Cauchy's

[1] Smyth, J.A.C.S. **50**, 536 (1928).
[2] Stranathan, Phys. Rev. **31**, 653 (1928).
[3] Williams and Krchma, J.A.C.S. **49**, 1676 (1927).

relation between the index of refraction and the wave-length, $n = A + B/\lambda^2$, it is possible to estimate the index of refraction for infinite wave-length. The electronic polarization can then be calculated from the Lorentz-Lorenz relation,

$$P_e = \frac{n^2 - 1}{n^2 + 2} \times \frac{M}{d}. \tag{3}$$

Since the total and the electronic polarizations are then known, the electric moment can be calculated provided that the atomic polarization is assumed to be zero and the electronic polarization remains constant. However, the atomic polarization is not known to be zero and the calculation of the electric moment on this basis gives values which are a little too high, the difference depending on the magnitude of the atomic polarization.

The total polarization may also be measured over as large a range of temperature as possible. Then by making use of relation (1) we get

$$P = P_0 + a/T$$

where $P_0 = P_e + P_a$ and $a = (4/9)\pi N u^2/k$. By plotting the total polarization times the absolute temperature against the temperature we get a straight line whose intercept on the vertical axis is a and whose slope is P_0. In this method the electric moment of the molecule is determined independently of the electronic and atomic polarizations provided these polarizations are not functions of the temperature. The sum of the atomic and electronic polarizations is also determined and if the electronic polarization is calculated the atomic polarization can be determined.

In the case of associating substances it is necessary to make measurements on the polarizations of dilute solutions of the polar substance in a non-polar solvent. These polarizations are then plotted against the concentration and the value for zero concentration is determined by extrapolation. The non-associating substances permit their polarizations to be measured in the pure state.

It was the purpose of this investigation to determine the electric moments and an estimate of the atomic polarization of a number of organic molecules. The molecules studied were, bromobenzene and iodobenzene in dilute solution with hexane, and p-xylene, m-xylene, and o-xylene as pure liquids. The bromobenzene and iodobenzene were chosen to complete the monohalogen substituted group since chlorobenzene has carefully been worked out by Smyth.[4] The molecules p-xylene, o-xylene and m-xylene were investigated to determine their electric moments by the method of temperature variation in order to find the effect of the position of the substituted radical. The molecules p-xylene, o-xylene and m-xylene are non-associating and can be investigated in the pure state.

Experimental

The method for determining the dielectric constants of the liquids and solutions is indicated in Fig. 1. The apparatus consisted essentially of a substitution bridge, a 476 kilocycle oscillator and a detector.

[4] Smyth and Morgan, J.A.C.S. **50**, 1547 (1928).

The oscillator.

The oscillator O consisted of one oscillatory circuit whose output was amplified by two tubes in parallel. The output of these tubes was in turn induc-

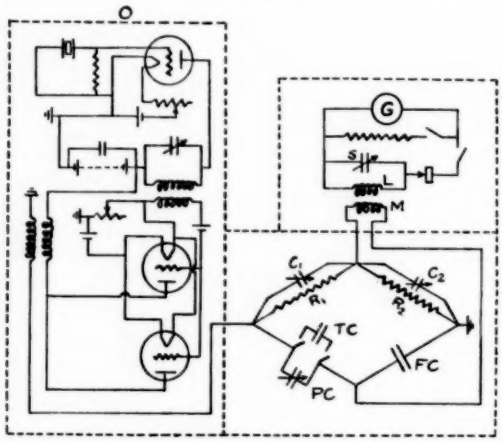


Fig. 1. A diagram of the circuit.

tively coupled to the bridge. The oscillatory circuit was maintained at a fixed frequency by means of a quartz crystal giving a frequency of 476 kilocycles.

The bridge.

The bridge consisted of two 1000 ohm resistances forming R_1 and R_2 one path of the bridge with a variable capacity C_1 and C_2 across each resistance. In the other path of the bridge there was a parallel set of condensers, consisting of the test condenser TC and a precision condenser PC, in series with a fixed capacity of 1400 mmf.

The detector.

The detector circuit consisted of a coil L inductively coupled to a coil M across the bridge. A condenser S connected in parallel with the coil L was adjusted to give parallel resonance for the frequency of 476 kilocycles. The point of balance was determined by means of a high sensitivity galvanometer and a carborundum crystal.

The test condenser.

The test condenser consisted of three concentric brass cylinders of diameters 3.0, 3.2, and 3.4 cm. The length of the outer two was 15 cm and of the inner 13 cm. The outer two were connected together by means of a circular brass ring soldering them at the top and bottom. The upper part of the outer cylinders was cut away so that there could be a free circulation of the cooling liquid around the condenser. The connection to the inner plate was made through a brass tube with a glass capillary as an insulator. The condenser was immersed in a cooling bath consisting of solid carbon dioxide and alcohol in a Dewar flask. The solutions were cooled to the lowest desired temperature

for the initial measurement and then the temperature was increased by means of a heating coil immersed in the bath. The temperatures were measured by means of a copper-constantin thermo-couple and a White potentiometer.

Measurement of densities.

The densities of the solutions and liquids were measured by means of a pycnometer similar to that used by Isnardi.[5] It was made of Pyrex and had a volume of 15 cubic centimeters.

The purity of substances.

Petroleum hexane from Eastman Kodak Company was used entirely for the non-polar solvent. Smyth has shown that in this type of work, there is no error in using the petroleum hexane in place of the pure normal hexane. The brombenzene, iodobenzene and the xylenes were chemically pure Eastman Kodak substances and no further effort was made for purification.

Method of procedure.

The bridge was first brought to a balance by adjusting the precision condenser and the condensers shunting the resistances. (The shunting condensers

TABLE I. *Interpolated values of the dielectric constants and densities* (d_4^t) *of solutions of bromobenzene and iodobenzene in hexane.*

	Bromobenzene					Iodobenzene				
Mol. fract.	0.04	0.08	0.12	0.16	0.20	0.02	0.04	0.06	0.08	0.10
t°C.										
−60	2.152 0.7652	2.295 0.7969	2.437 0.8283	2.580 0.8603	2.722 0.8923	2.080 0.7560	2.149 0.7785	2.217 0.8010	2.288 0.8235	2.356 0.8460
−50	2.132 0.7561	2.270 0.7877	2.408 0.8190	2.546 0.8509	2.683 0.8828	2.068 0.7471	2.130 0.7693	2.397 0.7920	2.263 0.8144	2.330 0.8368
−40	2.113 0.7471	2.246 0.7785	2.379 0.8098	2.512 0.8415	2.644 0.8732	2.046 0.7383	2.110 0.7601	2.175 0.7830	2.240 0.8055	2.303 0.8276
−30	2.094 0.7382	2.222 0.7693	2.350 0.8004	2.478 0.8320	2.606 0.8638	2.028 0.7289	2.090 0.7509	2.153 0.7740	2.216 0.7961	2.277 0.8183
−20	2.075 0.7290	2.198 0.7600	2.322 0.9711	2.445 0.8227	2.568 0.8543	2.010 0.7200	2.072 0.7419	2.133 0.7649	2.194 0.7869	2.253 0.8090
−10	2.056 0.7200	2.174 0.7510	2.294 0.7819	2.412 0.8133	2.530 0.8447	1.992 0.7105	2.053 0.7330	2.112 0.7557	2.170 0.7777	2.229 0.7996
0	2.037 0.7112	2.150 0.7418	2.265 0.7727	2.379 0.8039	2.492 0.8352	1.957 0.7015	2.033 0.7240	2.090 0.7465	2.147 0.7684	2.402 0.7902
10	2.018 0.7020	2.126 0.7327	2.237 0.7635	2.345 0.7946	2.454 0.8258	1.959 0.6928	2.015 0.7151	2.070 0.7372	2.123 0.7591	2.177 0.7808
20	1.999 0.6931	2.102 0.7235	2.208 0.7542	2.312 0.7850	2.416 0.8162	1.943 0.6838	1.998 0.7062	2.047 0.7280	2.100 0.7497	2.153 0.7714
30	1.980 0.6842	2.078 0.7143	2.180 0.7450	2.279 0.7757	2.378 0.8066	1.928 0.6749	1.978 0.6972	2.027 0.7187	2.077 0.7403	2.128 0.7617
40	1.961 0.6752	2.054 0.7051	2.151 0.7356	2.245 0.7663	2.339 0.7970	1.912 0.6666	1.960 0.6882	2.009 0.7095	2.057 0.7308	2.106 0.7523
50	1.942 0.6662	2.030 0.6960	2.122 0.7262	2.211 0.7568	2.300 0.7875	1.896 0.6576	1.944 0.6792	1.992 0.7002	2.038 0.7214	2.085 0.7427
60	1.923 0.6578	2.006 0.6868	2.093 0.7168	2.177 0.7473	2.261 0.7780	1.881 0.6488	1.927 0.6700	1.973 0.6910	2.016 0.7120	2.061 0.7330

Infinite dilution polarizations of C_6H_5Br and C_6H_5I

	−60	−50	−40	−30	−20	−10	0	10	20	30	40	50	60
C_6H_5I	98.7	96.8	95.3	93.8	92.5	91.2	89.9	88.1	86.5	84.5	82.6	81.3	80.0
C_6H_5Br	100.0	99.0	97.7	96.2	94.4	92.8	91.3	90.0	88.9	87.7	86.2	85.0	84.0

[5] Isnardi, Zeits. f. Physik 9, 153 (1922).

were necessary since the resistances could not be made entirely non-inductive and non-capacitive and since there are always stray capacity effects.) After this balance was obtained, the test condenser containing the substance whose dielectric constant was to be determined was connected in parallel with the precision condenser. A new balance was then determined by adjusting *only* the precision condenser. The change in the setting of the precision gave the capacity of the test condenser. The correction for the leads and the spacers of the test condenser was determined by filling the condenser with benzene, $e = 2.273$ at 25°C.[4,6,7] The capacity of the test condenser after the correction was made was found to be 217.1 mmf. On account of the low conductance of the substances under investigation, it was never necessary to make any adjustment of the shunting condensers to secure a balance. The probable error in the determination of the dielectric constants was about 0.2 percent.

Results

The experimental values, after interpolation, of the dielectric constants and densities for the low concentrations (0 to 0.2 mols) of bromobenzene and iodobenzene in hexane were found to decrease linearly with the temperature

Table II. *Characteristic constants of the substances, p-xylene, o-xylene, m-xylene and hexane.*

t°C	Dielectric constant	Density d_4^0	Polarization	t°C	Dielectric constant	Density d_4^0	Polarization
	p-xylene				hexane		
20	2.264	0.8611	36.51	−60	2.010	0.7332	29.58
30	2.249	0.8522	36.59	−50	1.995	0.7244	29.60
40	2.234	0.8433	36.66	−40	1.980	0.7157	29.62
50	2.219	0.8345	36.72	−30	1.966	0.7069	29.66
60	2.024	0.8257	36.79	−20	1.952	0.6982	29.70
70	2.189	0.8169	36.85	−10	1.937	0.6895	29.73
80	2.175	0.8080	36.94	0	1.923	0.6807	29.76
90	2.160	0.7990	37.02	10	1.908	0.6719	29.77
100	2.145	0.7900	37.09	20	1.893	0.6631	29.79
110	2.130	0.7810	37.16	30	1.878	0.6543	29.81
120	2.115	0.7720	37.23	40	1.864	0.6454	29.82
130	2.100	0.7630	37.30	50	1.850	0.6366	29.86
				60	1.836	0.6278	29.89
	o-xylene				m-xylene		
−20	2.701	0.9124	42.06	−40	2.482	0.9145	38.35
−10	2.674	0.9039	42.03	−30	2.462	0.9063	38.35
0	2.648	0.8955	42.00	−20	2.444	9.8978	38.39
10	2.621	0.8872	41.94	−10	2.426	0.8894	38.42
20	2.594	0.8793	41.85	0	2.406	0.8807	38.43
30	2.567	0.8706	41.80	10	2.386	0.8723	38.44
40	2.541	0.8623	41.74	20	2.367	0.8639	38.43
50	2.514	0.8541	41.65	30	2.348	0.8556	38.43
60	2.488	0.8458	41.58	40	2.329	0.8472	38.44
70	2.461	0.8376	41.47	50	2.309	0.8386	38.42
80	2.434	0.8296	41.35	60	2.290	0.8304	38.41
90	2.407	0.8216	41.22	70	2.271	0.8219	38.40
100	2.381	0.8132	41.12	80	2.251	0.8133	38.38
110	2.354	0.8048	40.99	90	2.322	0.8051	38.35
120	2.328	0.7966	40.86	100	2.212	0.7966	38.31
130	2.302	0.7866	40.71	110	2.193	0.7881	38.29
				120	2.174	0.7999	38.25

[6] Hartshorn and Oliver, Proc. Roy. Soc. 664 (1929).

[7] Errera, Journ. de Physique 304 (1929).

and are given in Table I. The experimental values of the dielectric constants of the pure non-associating substances hexane, p-xylene, o-xylene and m-xylene are given in Table II. The relation of the dielectric constant and tem-

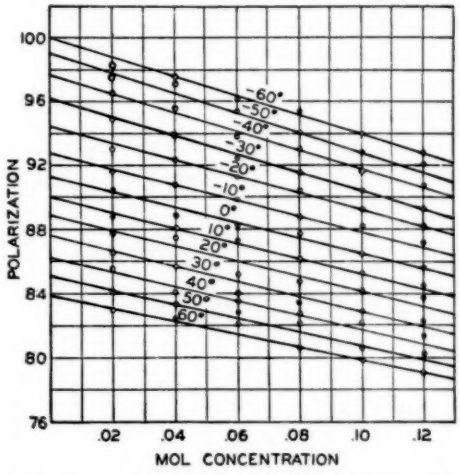


Fig. 2. The extrapolation to infinite dilution of the polarization of bromobenzene

perature is a linear one. The densities of the xylenes were also found to be linear functions of the temperature.

The total polarizations of the binary mixtures, bromobenzene and iodobenzene in hexane, were calculated from the Clausius-Mossotti relation (2)

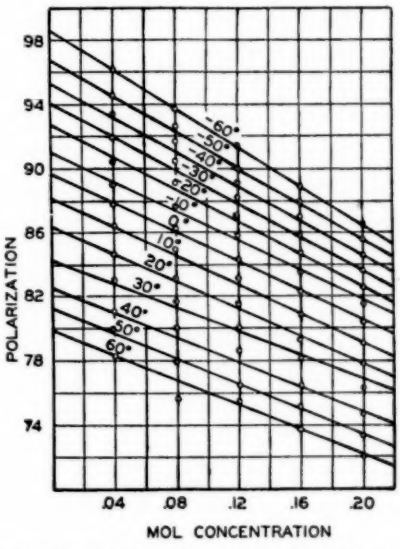


Fig. 3. The extrapolation to infinite dilution of the polarization of iodobenzene.

using the effective molecular weight of the mixture. The polarization due to the polar component was then calculated and plotted against the concentration of the polar substance as shown in Figs. 2 and 3. These curves were extra-

polated to zero concentration (infinite dilution) and the values $P_{2\infty}$ determined. The polarization increases to a maximum at zero concentration. These extrapolated values of the polarizations are the ones that should satisfy the Debye relation (1).

The total polarizations of the m-o-p-xylenes can be determined from measurements made in the pure state since for these substances Williams has shown that the polarizations of the polar molecule in solution with a non-polar substance is directly proportional to the concentration and they are therefore non-associating.[3]

Discussion of Results

Comparison with other data.

Table III gives a comparison of the results obtained on the density and dielectric constants with the corresponding results of other observers.

Table III.

Substance	Temperature	Density $d^{t}_{4^0}$	Dielectric constant	Observers
Hexane	60°C	0.6473	1.843	Smyth and Morgan[4]
Hexane	60°C	0.6430	1.856	Morgan and Lowry[9]
Hexane	60°C	0.6278	1.836	Author
p-xylene	25°C	0.8567	2.265	Williams[3]
p-xylene	17°C		2.259	I.C.T.
p-xylene	25°C	0.8545		I.C.T.
p-xylene	25°C	0.8549	2.257	Author
o-xylene	25°C	0.8764	2.507	Williams[3]
o-xylene	25°C	0.8754	2.58	I.C.T.
o-xylene	25°C	0.8749	2.580	Author
m-xylene	20°C		2.392	Isnardi[5]
m-xylene	20°C		2.374	I.C.T
m-xylene	20°C	0.8639	2.367	Author
m-xylene	20°C	0.8642		Landolt-Bornstein

Table IV gives the values of the electric moments and the atomic polarizations of the substances investigated. The value of the moment of bromobenzene has been determined by various observers,[8,10,11,12,13] who have ob-

Table IV.

	Electric moment $\times 10^{18}$	Atomic polarization
Hexane C_6H_{14}	0	0.6
p-xylene C_8H_{10}	0 (assumed)	2.3
m-xylene C_8H_{10}	0.34	4.0
o-xylene C_8H_{10}	0.44	4.5
Bromobenzene C_6H_5Br	1.35	13.9
Iodobenzene C_6H_5I	1.25	18.3

[8] Williams and Allgier, J.A.C.S. **49,** 2416 (1927).
[9] Morgan and Lowry, Journ. of Phys. Chem. **34,** 2385 (1930).
[10] Das and Roy, Ind. Journ. of Phys. **15,** 441 (1930).
[11] Hojendahl, Phys. Zeits. **30,** 391 (1928).
[12] Bergmann, Engel, Sandor, Zeits. f. Phys. Chem. **10** Abt. B p. 106.
[13] Muller and Sack, Phys. Zeits. **31,** 518 (1930).

tained values ranging from 1.49×10^{-18} to 1.56×10^{-18}. All these observations, except that of Das and Roy, were made by the single temperature method. Das and Roy studied bromobenzene in dilute solution with carbon tetrachloride over a temperature range from 10°C to 60°C. Their extrapolated values of the polarization agree with those of the author the difference in the value of the moment determined is due to a lower value of the polarization at temperatures lower than those studied by Das and Roy. The value obtained for the moment of iodobenzene agrees with that obtained by other observers, 1.25×10^{-18} by Walden and Werner,[14] and 1.30×10^{-18} by Bergmann Engel and Sandor,[12] who used the single temperature method.

Effect of different substituents.

The effect of substituting chlorine, bromine and iodine in three different types of molecules is shown in Table V. From this table it is seen that in

TABLE V.

Substituent X	Electric moment $\times10^{18}$ C_6H_5X	CH_3X	HX
Chlorine	1.52[4]	1.56[9]	1.04[15]
Bromine	1.35 (author)	1.45[9]	0.79[15]
Iodine	1.25 (author)	1.35[9]	0.38[15]

these three series of compounds, C_6H_5X, CH_3X and HX, an increase in the atomic weight of the substituted atom produces a decrease in the electric moment of the molecule. The results obtained for bromobenzene (C_6H_5Br) and iodobenzene (C_6H_5I) are thus in agreement with this effect.

The induced polarization a function of the temperature.

Table II shows an increase in the total polarization with the temperature for the case of p-xylene. Since p-xylene is a symmetrical molecule it should have no electric moment, and its polarization, the sum of the electronic and atomic polarizations, should remain constant. The increase of polarization means that the induced polarization increases with temperature. This increase of the induced polarization with the temperature, in the case of p-xylene, leads to a negative intercept as shown in Fig. 4. The intercepts for the m-xylene and o-xylene molecules correspond to moments of zero and 0.34×10^{-18} respectively. However, since p-xylene is a symmetrical molecule, if the three straight lines in Fig. 4 representing the xylenes, are shifted, so the one representing p-xylene passes through the origin, intercepts corresponding to 0.34×10^{-18} and 0.44×10^{-18} as values of moments are obtained for the m-xylene and o-xylene molecules respectively. These results can be interpreted in the following manner; the total polarization of p-xylene increases with the temperature since the induced polarization increases and there is no dipole polarization; in the case of m-xylene the total polarization (see Table II) remains practically constant, since the increase in the induced

[14] Walden and Werner, Zeits. f. physik. Chemie **12**, Abt. B (1930).
[15] Zahn, Phys. Rev. **24**, 400 (1924).

polarization is just overcome by the small dipole polarization; for o-xylene this dipole polarization is still more prominent and it produces a decrease in the total polarization. Smyth[16] has found the same effect in the case of the

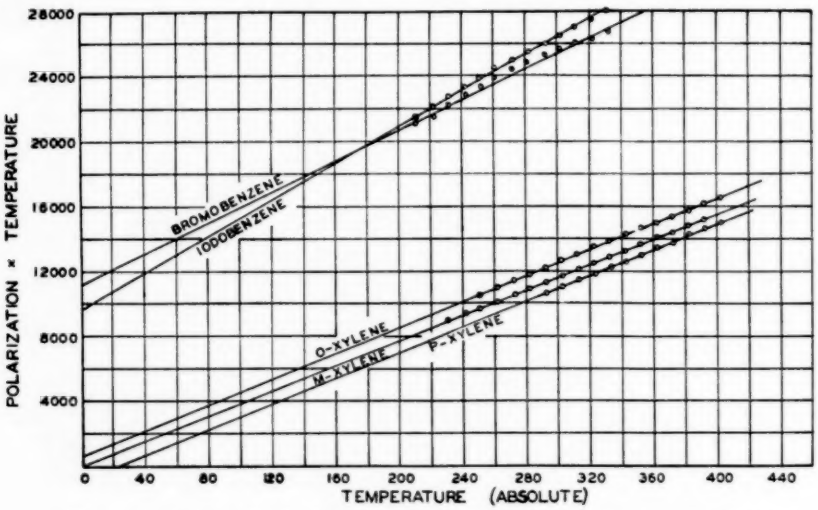


Fig. 4. Relation of polarization times temperature vs. temperature for bromobenzene, iodobenzene, p-m-o-xylenes.

normal non-polar paraffins, the total polarization increasing with the temperature at a rate of 1 percent per 100 degrees.

Summary

1. The dielectric constants and densities of dilute solutions of bromobenzene and iodobenzene in hexane have been measured. From these measurements the values of the electric moments and the atomic polarizations have been determined for the bromobenzene and iodobenzene molecules.

2. The dielectric constants and densities of pure p-xylene, m-xylene, o-xylene and hexane have been determined and the electric moments and atomic polarizations of these molecules calculated.

3. Evidence has been given to show that, as the atomic weight of a substituted atom is increased, the value of the electric moment of the molecule is decreased. It has been pointed out that the results of this investigation give additional evidence of this rule.

4. Evidence has been deduced which suggests that the induced polarization is a function of the temperature and that this effect may neutralize or mask the effect of a small dipole.

The author wishes to express his appreciation of the advice and criticism given by Dr. Alpheus Smith and Dr. C. P. Smyth during this investigation.

[16] Smyth and Stoops, J.A.C.S. **52**, 3546 (1930).

ON THE STABILITY OF CERTAIN HETEROPOLAR CRYSTALS

By H. M. Evjen*
California Institute of Technology

(Received December 14, 1931)

Abstract

On the basis of the conventional model of heteropolar crystals it is shown that an irregular crystalline behaviour is to be expected in the region of small values of the exponent, p, of the repulsive energy. In the course of this demonstration Madelung's method of obtaining the Coulomb potential for the crystals of the cubical system is justified, and a new, very simple method of calculating the Madelung constant is described. This method is applied to the three crystals of the cubical system and other applications are suggested. The irregular behavior of crystals in the region of small values of the repulsive exponent is shown to be manifested as an instability against various variations by which the geometry of the lattice is altered. In particular, the instability is demonstrated in the case of the "calcite family" of crystals which is evolved by a continuous process from a single parameter ϕ. The NaCl- and CsCl-types of crystals are members of this family, and a continuous mode of transition from one to the other is thus available. These two types of crystals, in general, are unstable against a variation in which the parameter ϕ is changed, (ϕ-variation), and for small values of the exponent, p, all members of the "calcite family" are shown to have a tendency to fall apart into one-dimensional crystals of the Madelung type. Another result of these considerations is that a skew structure, such as that of calcite may be accounted for on the basis of purely central forces. Finally the regions in which respectively one-dimensional and two-dimensional crystals, and the crystals of the cubical system are most stable are calculated. The bearing of these results on the theory of the secondary structure is briefly discussed.

I. Introduction

IN THE following we shall assume that the potential energy of a molecule may be represented thus:

$$U_m = -(ne)^2/r + A/r^p + U_p. \tag{1}$$

The first term is the Coulomb energy of two point charges, $\pm(ne)$, situated a distance r apart; U_p is the energy of polarization which arises as a result of the deformability of the ions, and A/r^p is the potential energy of the repulsive forces between the ions.

This expression for the potential energy is at best a first approximation. In the first place, the term purporting to represent the Coulomb energy, in general, is merely the first member in a power series expansion to which should be added also terms of the nature of homopolar bindings. As we shall see in the course of this discussion, these higher order terms cannot, in general, be neglected. In the second place, the representation of the energy of repulsion is not functionally correct.[1] As redeeming features we have the two

* National Research Fellow.

[1] See, for instance, J. C. Slater, Phys. Rev. 23, 488 (1924) or G. A. Tomlinson, Phil. Mag. 11, 1009 (1931).

arbitrary parameters A and p which may be adjusted empirically in such a way as to make our approximate functional representation of the energy closely approach the true representation over the small interval of the independent variable, r, in which we are primarily interested.

The simple hypothesis (1), in the hands of Born,[2] Madelung,[3] and others, has led to fruitful results insofar as the socalled structure insensitive properties of heteropolar crystals have been satisfactorily accounted for. There are reasons for believing[4] that the complete failure of the theory to explain the structure sensitive properties of such crystals is not primarily due to the approximate nature of the fundamental assumption (1), but is due to the presence of a regular secondary structure within the crystal, predicted by Zwicky,[5] and later verified experimentally.[6] The existence of a secondary structure appears to be intimately connected with the existence of the crystal itself,[7] and therefore appears to be, within wide limits, independent of the particular nature of the function by which we choose to represent the potential energy. In particular, it has been shown that secondary structures of various types[8] are not inconsistent with energy functions of the type (1). From this we are led to believe that small inaccuracies in the representation (1) should lead to correspondingly small inaccuracies in any calculated property of the crystal, (as exemplified by the comparatively slight spread between calculated and observed structure insensitive properties); that is to say, if proper account is taken of the secondary structure. This is our justification for considering the simple energy function (1) adequate for our present needs.

II. The Potential Energy of an Ideal Crystal

In the ideal models with which here we shall be concerned the lattice points are symmetry centers at which the electric vector vanishes. Hence the energy of polarization is zero, and the potential energy per molecule may be represented as,

$$U = -M(ne)^2/\rho + RA/\rho^p \tag{2}$$

where M, which we shall call Madelung's constant, is a summation over the lattice points and depends only on the geometry of the crystal; R similarly represents a summation over the lattice points and depends also on the exponent of the energy of repulsion, p. Eq. (2) involves the assumption that the three different combinations of ions which occur in a crystal are identical in regard to the mutual repulsive forces. This naturally will derogate the

[2] M. Born, Atomtheorie d. Festen Zustandes, Teubner, 1923; Enz. d. Math. Naturwiss. V3, 4.

[3] E. Madelung, Phys. Zeits. **19**, 524 (1918).

[4] F. Zwicky, Helv. Physica Acta **3**, 269 (1930); and **4**, 49 (1931).

[5] F. Zwicky, Proc. Nat. Acad. Sci. **15**, 816 (1929).

[6] A. Goetz, Proc. Nat. Acad. Sci. **16**, 99 (1930); F. Zwicky, Proc. Nat. Acad. Sci. **16**, 211 (1930); M. Straumanis, Zeits. f. physik. Chem. (B) **13**, 316 (1931); H. Johnson, Journ. Franklin Inst. Oct. (1931). F. Bitter, Phys. Rev. **38**, 1903 (1931).

[7] F. Zwicky, Proc. Nat. Acad. Sci. Sept. (1931).

[8] See references 4 and 5, and also, H. M. Evjen, Phys. Rev. **34**, 1385 (1929).

generality of our results, without, however, seriously impairing their qualitative value. By the usual method we stabilize our model against uniform compression by letting,

$$\frac{\partial U}{\partial \rho} = 0. \tag{3}$$

It is easily seen that for p greater than unity $\partial^2 U/\partial \rho^2$ is always greater than zero, so that the extreme obtained by Eq. (3) is always a minimum.

If we substitute the value $\rho = d$ obtained from Eq. (3) into Eq. (2) we obtain,

$$U = -\left[(p-1)/p\right] \cdot \left[M(ne)^2/d\right] \tag{4}$$

which is always negative, so the crystal is always stable against all variations by which it would fall completely apart. Eq. (4) also may be written,

$$U = -f\left(\frac{M}{R^{1/p}}\right)^{p/p-1} \tag{5}$$

where $f = (p/p-1)\left[(ne)^{2p}/pA\right]^{1/p-1}$ is independent of the geometry of the crystal and depends only on the physical properties of the ions involved. The ratio $M/R^{1/p}$ then is a measure of the stability of the crystal.[9] The larger this ratio the more stable the crystal. This ratio we shall call the *stability factor* of the crystal.* The inaccuracies of the hypothesis (1), to which we have called attention, of course detract from the reliability of the stability factor. In particular, Pauling[10] has shown that it is not absolutely reliable in comparing the stability of crystals of different geometrical complexions. In view of Hund's results, however, it may at least be considered as a qualitative guide to the stability of crystals.

The static model obtained by this procedure, as we have seen, is always stable against uniform compression or expansion. The question remains whether it is stable against variations of a different type. In particular, is it stable against variations by which the geometry of the lattice is altered? This has been generally taken for granted. Recently it has been shown, however, that this is not, in general, true. Zwicky[11] has shown that for p less than 6, (5.95 to be more exact) the NaCl-type of crystal is unstable against a "residual-ray variation." The writer[12] similarly has found several variations against which the same type of crystal is unstable; for instance, Poisson's ratio becomes negative for p less than 6.2 and the crystal there becomes unstable against elongation and contraction. Thus there is a considerable variety of variations against which the ideal crystal is unstable in the region of small exponents, p. In the following we shall point out a fact which may be the fundamental cause of this instability.

[9] F. Hund, Zeits. f. Physik **34**, 833 (1925).
[10] L. Pauling, Zeits. f. Krist. **69**, 35 (1928).
[11] F. Zwicky, Phys. Rev. **38**, 1772 (1931).
[12] To be published shortly.

* Not to be confused with the dynamic stability factor given by the normal modes of vibration of the system.

III. The Coulomb Energy of an Ideal Crystal

In accordance with Eq. (1) the Coulomb energy of an ideal crystal may be written as follows:

$$U_e = -M(ne)^2/d$$

where

$$M = \sum_{\alpha} \sum_{\beta} \sum_{\gamma} \frac{(-1)^{\Phi(\alpha,\beta,\gamma)}}{\frac{r(\alpha,\beta,\gamma)}{d}}. \tag{7}$$

Several ingenious methods[13] have been devised for evaluating the sum of the series (7). Madelung's method depends for its success and simplicity on the peculiar order in which the potentials of the individual ions are summed. In fact, the ions are collected into groups, each group forming a linear element within the crystal. The potential of one such linear element is calculated, and the potential of the crystal as a whole is then said to be the sum of the potentials due to these linear elements. Such a procedure, in general, is not strictly correct. It is only permissible when the series in question is such that its sum is independent of the order of summation. Although it is not at all obvious, Madelung's method, as we shall see, is justified in the case of the crystals of the cubical system. If, however, his method is applied indiscriminately to a calculation such as that of the change in energy by a "residual-ray variation" the result, in general, will be incorrect, since the order of summation here is not immaterial, and the result of the calculation will depend on the external shape of the region affected by the variation. These difficulties are avoided, and incidentally a new and very simple method of calculating Madelung's constant is obtained, if we remember that the building stones of the crystal are not the individual ions but rather the elementary cells of the crystal. In other words, we propose to sum the potentials of the individual cells rather than those of the individual ions; that is,

$$U_e = +(ne) \sum_{l,m,n} \Psi(l, m, n) \tag{8}$$

where $\Psi(l, m, n)$ is the potential at the lattice point (0, 0, 0) of the cell (l, m, n). We now collect the terms of the series (8) in order of increasing distance of the corresponding cells from the origin. It is then easy to show that the remainder after summing N of these terms approaches the integral,

$$Q = (ne) \cdot \sigma \int\int\int \psi(x, y, z) dx dy dz \tag{9}$$

where σ is the number of elementary cells per unit volume, and the integration is to be extended over the volume outside the sphere of radius $R(N)$ to which the direct summation has been carried. In the case of the crystals of the cubical system we may choose for the elementary cell the elementary cube, assigning the weights 1, 1/2, 1/4, or 1/8 to the charges according as they are

[13] See reference 2 for summary.

situated in the interior, on the face, on the edge, or on the corner of the cube. The total charge of the volume cell then, of course, is zero. The potential of the volume cell, referred for instance to its body center, now may be found by an expansion in terms of Legendre polynomials.[14] If r be the radius vector from the origin to the center of the volume cell, and l_i the radius vector from the center of the volume cell to the i-th ion, e_i, of this cell, we have, in fact:

$$\psi = \sum_{n=0}^{\infty} \frac{A_n}{r^{n+1}} \tag{10}$$

$$A_n = \sum_i e_i l_i^n P_n(\cos\theta_i) \tag{11}$$

and

$$\begin{aligned} \cos\theta_i &= \left(\frac{x}{r}\right)\xi_i/l_i + \left(\frac{y}{r}\right)\eta_i/l_i + \left(\frac{z}{r}\right)\zeta_i/l_i \\ &= (a+b+c)/l_i \end{aligned} \tag{12}$$

where (x, y, z) are the coordinates of the center of the volume cell, and (ξ_i, η_i, ζ_i) the coordinates of the i-th ion with respect to the translated Cartesian coordinate system having its origin at the center of the volume cell. Consequently,

$$\begin{aligned} A_n = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{n!} \sum_i e_i \Big[&(a+b+c)^n \\ &- \frac{n(n-1)}{2(2n-1)} l_i^2 (a+b+c)^{n-2} + \text{etc.} \Big] \end{aligned} \tag{13}$$

A_n therefore will depend linearly on terms of the type $(x/r)^\alpha \cdot (y/r)^\beta \cdot (z/r)^\gamma$, where $\alpha+\beta+\gamma = n-2s$. It follows that, if we form the integral over the surface of a sphere:

$$B_n = \int_0^{4\pi} A_n d\Omega$$

$B_{2n+1} = 0$. It is therefore sufficient to consider the terms where n, α, β, and γ are even. By a slight transformation of Eq. (13) we obtain:

$$A_{2n} = \sum_{t=0}^{n} L_t^{2n} F_t \tag{14}$$

where L_t^{2n} denotes the t-th coefficient in the Legendre polynomial of degree $2n$.

$$F_t = \sum_{\Gamma} \frac{2(n-t)!}{2p!\ 2q!\ 2s!} M(p, q, s, t) \left(\frac{x}{r}\right)^{2p} \left(\frac{y}{r}\right)^{2q} \left(\frac{z}{r}\right)^{2s} \tag{15}$$

[14] See, for instance, M. Mason and W. Weaver, The Electromagnetic Field, p. 15.

where Γ denotes a summation over all the sets of integers $p+q+s=n-t$, and $M(p, q, s, t)$ is a $2n$-th order electric moment of the elementary cell, given by:

$$M(p, q, s, t) = \sum_i e_i l_i^{2t} \xi_i^{2p} \eta_i^{2q} \zeta_i^{2s}. \tag{16}$$

For the three types of crystals of the cubical system we have:

A. NaCl-type (Fig. 1)

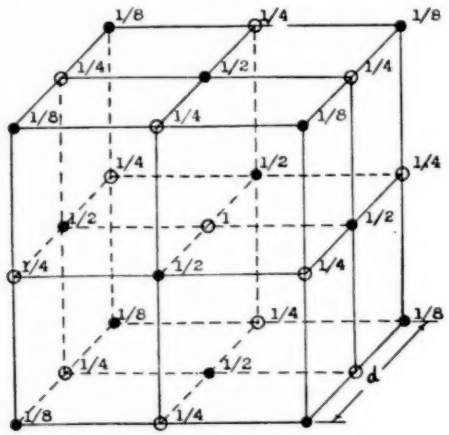


Fig. 1. Elementary cell of the NaCl-type.

$$M(p, q, s, t) = k_1 e(d/2)^{2n} \tag{17}$$

where $k_1 = -3^t + 2^t(0^p + 0^q + 0^s - (0^{pq} + 0^{qs} + 0^{sp})$, and the convention has been adopted:

$$0^m = \begin{cases} 0 \text{ if } m \neq 0 \\ 1 \text{ if } m = 0. \end{cases}$$

Here it is easily seen that $A_{2s+1}=0$, and $A_0=A_2=A_4=0$, so that the first term in the expansion of the potential of the element of an NaCl crystal goes as the inverse seventh power of the distance. This is a significant result since it shows that for a heteropolar crystal of this type the forces of cohesion act almost exclusively between neighboring atoms. In fact, the elementary cube of which an ion is a central member contributes approximately 84 percent to the Coulomb energy of that ion. This may be seen by direct summation with reference to Fig. 1. The contribution of the first elementary cube to the energy of the central ion is:

$$\begin{aligned} U_1 &= [-(6/1)(\tfrac{1}{2}) + (12/2^{1/2})(\tfrac{1}{4}) - (8/3^{1/2})(\tfrac{1}{8})]e^2/(d/2) \\ &= -2.92e^2/d \end{aligned}$$

as against the total energy as calculated by Madelung:

$$U_c = -3.496e^2/d; \ (M_1 = 3.496).$$

This immediately suggests a simple method of calculating the Madelung constant: Sum directly the potentials of the ions over a cube of side nd

around the origin. The potentials of the ions inside the cube are summed in the ordinary manner, each being given unit weight; the potentials of the ions forming the surface of the cube, however, are given the weights 1/2, 1/4, or 1/8 according as they are situated on a face, an edge, or a corner of the cube. This procedure, of course, is equivalent to summing by cells rather than by ions. If n be increased by unit steps, that is, by adding cubical shells, an extremely rapidly converging series is obtained, which moreover alternates. Narrow limits are therefore obtained between which the Madelung constant must lie. In this way the Madelung constant for the NaCl-type of crystal is obtained to three decimal places by summing over a cube of side not more than three or four times the lattice constant. This must be considered as more than sufficiently close in view of the inaccuracies of the fundamental assumption (1).

B. CsCl-type (Fig. 2)

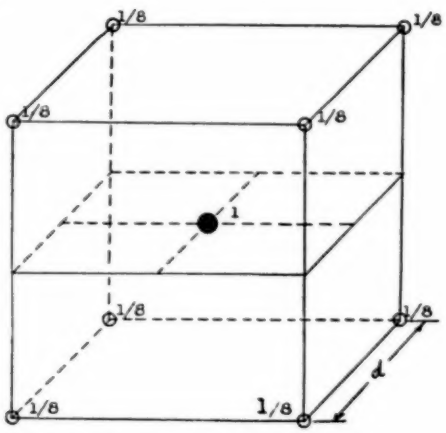


Fig. 2. Elementary cell of the CsCl-type.

For this type of crystal, we have,

$$M(p, q, s, t) = 3^t e(d/2)^{2n}. \tag{18}$$

From Eqs. (14) and (15) we therefore obtain $A_0 = A_2 = 0$, and $B_4 = \int_0^{4\pi} A_4 d\Omega = 0$, so that again the first term in the expansion of the potential of a spherical shell of elementary cells goes as the inverse fifth power of the radius of the shell. If one attempts to evaluate Madelung's constant for this type of crystal by the method outlined above, some difficulty is encountered due to the fact that two different limits seem to be approached according as one stops with a surface containing only positive ions or one containing only negative ions. The two limits, in fact, differ by an amount equal to the potential of a double-layer. This is to be expected, and is due to the fact that we sum over a cube rather than over a sphere. The result for a sphere, in the limit, obviously will be the average between the two limits for a cube. By taking this average after summing 12 distinct terms one obtains:

$$M_2 = 2.034$$

C. ZnS-type (Fig. 3)

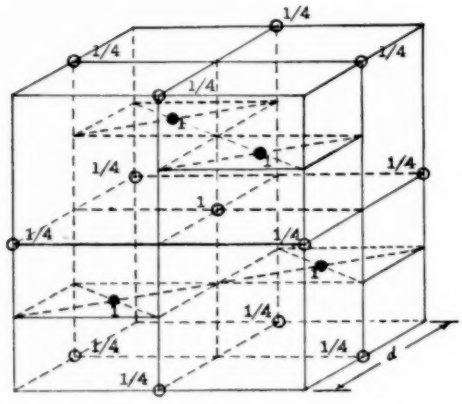


Fig. 3. Elementary cell of the ZnS-type.

The $2n$-th order electrical moment of the elementary cell here is given by:

$$M(p, q, s, t) = k_2 e(d/2)^{2n} \tag{19}$$

where

$$k_2 = 2^t(0^p + 0^q + 0^s) - 4 \cdot 3^t\left(\tfrac{1}{2}\right)^{2n}$$

Eqs. (14) and (15) therefore again yield $A_0 = A_2 = B_4 = 0$, and proceeding as before, one obtains:

$$M_3 = 3.78.$$

As a result of these considerations we can say that the series (8) is absolutely convergent and that its sum is independent of the order of summation for all the three types of crystals here involved. Madelung's method of summing therefore is justified, since his grouping of ions into linear elements may be obtained by a suitable grouping of the elementary cells. In general, the requirement for absolute convergence is $A_0 = A_1 = A_2 = 0$. The first coefficient, A_0, is always zero for electrically neutral matter. The second coefficient, A_1, always can be made to vanish by choosing the point of reference at the center of gravity of the charges of the elementary cell. Thus the most powerful restriction is the requirement that A_2 shall vanish. If A_2 is not zero we may get a conditional convergence if $B_2 = 0$. In general this would mean that the energy of the crystal would depend on its external macroscopic shape, which again might give rise to a type of secondary structure.

The simple method of calculating Madelung's constant which we have pointed out above, needless to say, is no longer of direct practical importance, since this constant has already been calculated for a great number of crystals with great accuracy.[15] The method, however, may be employed in the evaluation of various derivatives of the Coulomb potential; for instance, it may be applied to advantage in the calculation of the elastic constants, and is particularly useful when various irregular variations are imposed on the crystal.[12]

[15] O. Emersleben, Phys. Zeits. **24**, 73, 97 (1923).

Otherwise the most interesting result of the preceding considerations is the fact, to which we have already called attention, that the first term in the expansion of the Coulomb potential of an element of a cubical crystal goes effectually as the inverse seventh power of the distance. This fact has some important consequences. In the first place, when the exponent, p, of the potential of the repulsive forces is less than seven, the sphere of action of the repulsive forces is greater than that of the Coulomb forces. Consequently we should expect irregularities in the behavior of the crystal at these low values of the exponent p, such as the instability against various variations to which we have already called attention in paragraph 2.

The exceedingly small sphere of action of the Coulomb forces as well as of the forces of repulsion moreover makes it difficult to understand the high degree of regularity over extended regions found by x-ray analysis. This fact

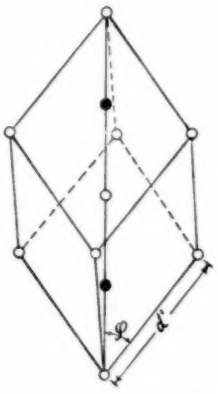


Fig. 4. Elementary cell of the "calcite family"

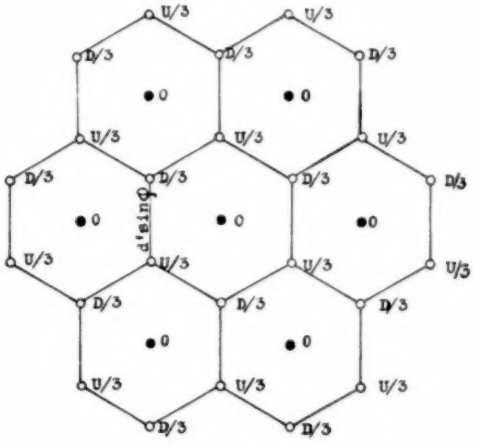


Fig. 5. Projective representation of the "calcite family"

seems to lend weight to Zwicky's contention[7] that really a different type of forces is responsible for the existence of crystals. Our consideration of the rapidity of convergence of the series involved may furnish clues as to how such forces can be brought into play, and how the necessary action over great distances can be effected. In the residual-ray variation[11] it is brought about by making the series involved conditionally convergent, but whether or not this is the only possible solution to the problem is a question to which we hope to return later.

The small radius of action of the Coulomb forces proper also elevates the importance of higher order terms in the expansion giving the Coulomb potential of an ion and also terms of the nature of homopolar bindings. Thus we are not in general justified in leaving out higher order terms in Eq. (1).

We shall now consider some more general evidence of their regular behavior of crystals for small values of the exponent p. This we shall first do in the case of a class of crystals which is evolved by a continuous process from a single parameter ϕ.

IV. Stability Factor of the Calcite Family of Crystals

The family of crystals obtained by assigning arbitrary values to the angle ϕ, (See Fig. 4), we shall for convenience call the calcite family of crystals. When the angle ϕ assumes a value approximately 45 degrees we obtain a model which, geometrically speaking, is a replica of calcite. For $\sin\phi = 1/3^{1/2}$ each ion in our model has eight next neighbors, and the model corresponds to the CsCl-type of crystal. On the other hand, for $\sin\phi = 1/3$ each ion has six next niehgbors, and the model becomes synonymous with the NaCl-type of crystal. In addition we obtain a crystal of the hexagonal type for $\tan\phi = 1/32^{1/2}$ with each ion having 12 next neighbors, a crystal, however,which does not appear to be of any interest for our present purposes.

A convenient representation of this family of crystals is given in Fig. 5. The points represent linear, mixed point successions of the Madelung type perpendicular to the plane of the paper. The letters and numerals indicate phase-difference. For instance, $oU/3$ indicates that the nearest positive ion is displaced up one-third of the distance, δ, between successive ions, which is given by:

$$\delta = \tfrac{3}{4}\cdot d'\cdot\cos\phi. \tag{20}$$

With reference to Fig. 5 we can immediately write down the Coulomb energy for this family of crystals, by making use of Madelung's method. It is,

$$U_c = U(0,0) + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}{}'[U_1(m,n) + U_2(m,n)] \tag{21}$$

where $U(0,0) = -8\log 2\ (ne)^2/3\cdot d'\cdot\cos\phi$
and $U_1(m,n) = -16\ (ne)^2/3\cdot d'\cdot\cos\phi$.

$$\cdot\sum_{t=0}^{\infty}K_0\left(\frac{2\pi}{3}\tan\phi(2t+1)(3m^2+(2n-m)^2)^{1/2}\right)\cos\frac{(2t+1)}{3}\pi$$

and $U_2(m,n) = 16\ (ne)^2/3\cdot d'\cdot\cos\phi$

$$\cdot\sum_{t=0}^{\infty}K_0\left(\frac{2(3)^{1/2}}{3}\pi\tan\phi(2t+1)(m^2+3(2n-m)^2)^{1/2}\right)\left[1+\cos\frac{(2t+1)}{3}\pi\right].$$

The energy of the repulsive forces may be most conveniently determined by direct summation. It is,

$$u = u_1 + u_2 \tag{22}$$

where

$$u_1 = 2^p A/(d'\cdot\cos\phi)^p$$

$$\cdot\sum_{t=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}[3t^2+(2m-t)^2+(3n+1)^2\cot^2\phi/4]^{-p/2}$$

and

$$u_2 = (4/3)^{p/2}[A/(d'\cdot\cos\phi)^p](S_1 - S_2)$$

where

$$S_1 = \sum_{t=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}{}'[t^2+3(2m-t)^2+n^2\cot^2\phi]^{-p/2}$$

and

$$S_2 = \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} [l^2 + 3(2m - l)^2 + (3n + 1)^2 \cot^2 \phi/12]^{-p/2}.$$

A sufficient accuracy here is attained by summing about ten distinct terms directly, after which the summation may be replaced by an integration. The modified Hankel functions $K_0(x)$ appearing in Eq. (21) will be found tabulated in the literature.[16] In conjunction with Eq. (2) the Eqs. (21) and (22) enable us to calculate the stability factor for this family of crystals. Points of contact with previously calculated results are obtained for the angles $\sin^{-1}(1/3^{1/2})$ and $\sin^{-1}(1/3)$.

In Fig. 6 we have plotted the stability factor for this family of crystals for three different values of the exponent p, namely, 5, 9, and infinity. The stability factor for an infinite exponent, of course, must be found by a limiting

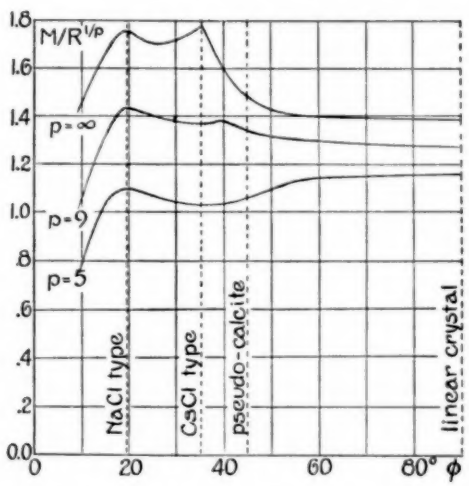


Fig. 6. Stability factor for the "calcite family."

process. For this exponent the repulsive effect of all but the next neighbors disappears, which is responsible for the discontinuity in the first derivative at arcsin (1/3) where the number of next neighbors changes from six to two.

For an infinite exponent the most stable members of this family crystals are seen to be the CsCl-type and the NaCl-type, in the order mentioned. As the exponent decreases the maximum in the neighborhood of the NaCl-type becomes more prominent than that in the neighborhood of the CsCl-type, and finally the NaCl-type becomes more stable than the CsCl-type. This is not the only effect of decreasing the exponent, however; a more startling effect is afforded by the shifting of the angles of maximum stability. The equilibrium position in the neighborhood of the NaCl-type of crystal is shifted only slightly, that in the neighborhood of the CsCl-type, on the other hand, quite appreciably, until finally the maximum occurs at an angle of 90

[16] E. Jahnke and F. Emde, Funktionentafeln, p. 135 (1928).

degrees which corresponds to a linear crystal. For still smaller exponents the linear crystal becomes even more stable than the NaCl-type.

In general, therefore, our models of the NaCl-type and the CsCl-type of crystals are unstable against the ϕ-variation. The instability of the NaCl-type would lead to a somewhat skewed structure, which, however, would be so slight as to make definite pronouncements hazardous in view of the inaccuracies of the fundamental assumption (1). The instability of the CsCl-type, on the other hand, is unmistakable, the maximum being shifted more and more toward the structure which we have called "pseudo-calcite". It is of interest to note that such a structure could be accounted for on the basis of purely central forces. Of interest is also the mechanism of transition, which here we have found, between the NaCl-type of crystal and the CsCl-type. The energy maximum separating the two types will be seen, in all cases, to be small compared to the total energy.

Finally attention is again called to the evidence of the irregular behaviour in the region of small values of the exponent p exhibited by Fig. 6. It will be noticed that all members of the calcite family of crystals are less stable than the linear crystal in this region. To further illuminate this point we have in the following extended Hund's theory[9] to include linear and planar crystals.

V. Extension of Hund's Theory

In the paper by Hund,[9] to which we have repeatedly referred, he calculates the stability factor for a number of different types of crystals. In view

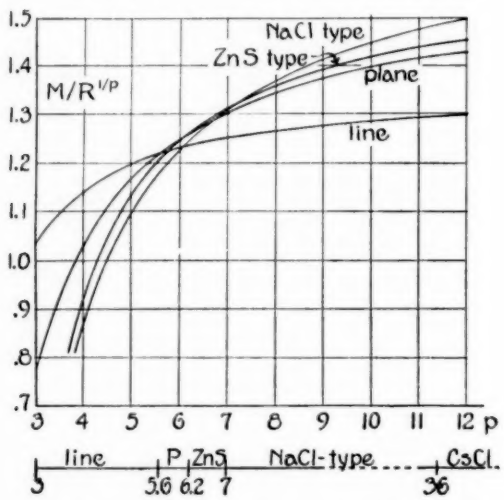


Fig. 7. Stability factor for the one-dimensional, two-dimensional, ZnS, and NaCl types of crystals.

of the tendency of our calcite family of crystals to fall apart into mixed linear point successions of the Madelung type for small values of the exponent p, it is of interest to extend Hund's work to include such linear crystals. Also, for the sake of generality, we shall include two-dimensional or planar crys-

tals, such as are obtained by isolating a 100-plane in the NaCl-type of crystal. The calculations closely parallel those of Hund and no further explanation is required except to say that wherever numerical series occur, the summation either has already been carried out in the literature, or it may be carried out to more than sufficient accuracy by direct summation of a few terms followed by the evaluation of an approximate integrated remainder.

Fig. 7 shows the result of these calculations together with the stability factors of the NaCl- and ZnS-types of crystals. The scale at the bottom of the figure indicates the regions in which each model is most stable. In spite of the idealized background of this diagram, its trend is so unmistakable that it seems permissible to draw some physical conclusions. The apparent instability of the ZnS-type, in particular, for an exponent five,[17] for instance, may not be real in view of the assumptions involved in Eq. (1), but it does not seem very probable that a modification of the energy function (1) could save the situation in general. In the various types of secondary structures, on the other hand, we have a source of energy which, in all cases, is ample to restore the stability of such a crystal as the one mentioned.[8,11]. Of these types of secondary structures the one based on a residual-ray variation[11] appears to be most suitable for the purpose of lining up the scale, Fig. 7, with observations, which is, however, a question to which we hope to return later.

The diagram 7 also appears to show some evidence in favor of the earlier type of secondary structure[8] which essentially consisted of a crystal within the crystal, that is, a regular network of two-dimensional crystals within the three-dimensional one. We are referring to the stability factor of the plane which will be seen to be nowhere very far below those of the three-dimensional crystals, and sometimes above. In order to establish the stability of this type of secondary structure it is, therefore, only necessary to devise a mechanism, such as the utilization of the energy of polarization, by which the comparatively small amount of energy separating the plane from the crystal may be gained without a corresponding loss. Such a mechanism has already been found by Zwicky,[8] and the success of his theory in accounting for the various structure sensitive properties is well known.

[17] Calculated on the basis of compressibility data. See, for instance, reference 2, p. 746.

OPERATING CHARACTERISTICS OF THE ELECTRO-OPTICAL SHUTTER

By Harold W. Washburn
University of California, Berkeley

(Received December 31, 1931)

Abstract

The electro-optical shutter is being employed at the University of California in the study of the electrical breakdown of gases and liquids. In these studies it is desirable to know the time it takes the shutter to close. A calculation of this time can be made from the electrical constants of the circuit and a knowledge of the rate at which the voltage drops across the spark gap. For some of the experimental conditions it is sufficiently accurate to base these calculations on an electrical circuit which replaces the actual distributed constants by lumped constants. In other cases however the error involved by this assumption is too great. It is the purpose of this paper to present an accurate solution of the electrical circuit taking into consideration that the constants are distributed, and by means of this solution to bring out the following important facts: (1) For relatively large distributed electrical capacities of the Kerr-cell leads the rate of closing of the shutter is greater than indicated by the lumped constant solution. (2) The rate of closing is materially increased by using leads separated only by a sheet of mica instead of spacing them farther apart in air. For completeness the results of a few experimental observations are also given and compared with results obtained by calculation.

Introduction

THIS paper is concerned with an accurate mathematical solution of the Abraham and Lemoine[1] type of electro-optical shutter. Until this solution was made we were using an approximate solution in which it was assumed that the electrical constants were lumped instead of distributed. The accurate solution which takes into consideration that the constants are distributed was originally made in order to determine the type and magnitude of errors involved in the approximate solution. However, it accomplished more than this. Calculations demonstrate the important fact that it is possible materially to speed up the rate of closing of the shutter by increasing the distributed capacity of the leads to the Kerr-cell. Moreover one is enabled to trace out the travelling waves of current and voltage and thereby obtain a true picture of the operation of the shutter.

Arrangement of Apparatus

A diagram of the arrangement of the apparatus is shown in Fig. 1. It will be noted that there is an optical system and an electrical circuit. In the optical system the two Nicol prisms are crossed and the direction of the electrical field in the Kerr-cell is placed at 45 degrees to the plane of polarization of the light passing through the first Nicol prism. This means that light can

[1] Abraham and Lemoine, C. R. **129**, 206 (1899).

pass from the spark gap through the prisms to the eye only when a voltage is on the Kerr-cell. The light from the spark which will be observed therefore is that light which passes through the Kerr-cell before the cell has discharged to an extent which effectively closes the shutter.

In the electrical system C_1 and R_1 are of such a size as to make the time it takes the voltage to reach its final value on the gap of the order of one second. The distributed resistance between C_2 and the spark gap is large enough to make the discharge of C_2 through the gap aperiodic. The distributed resistance between the spark gap and Kerr-cell is made large enough so that the second and all succeeding oscillations of voltage across the Kerr-cell are so small that the light transmitted is too feeble to be observed.

Operation of the Kerr-Cell Circuit

The solution of the electrical circuit shown in Fig. 1 is extremely complicated but fortunately the arrangement of the experimental set-up allows an

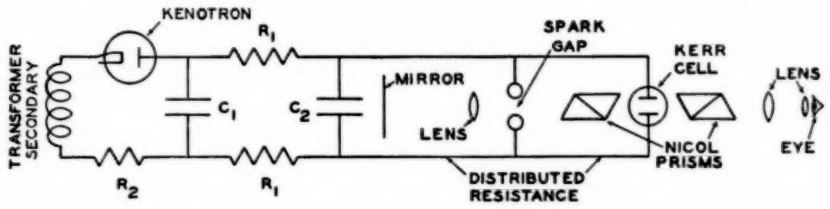


Fig. 1. Diagram of electrical and optical circuits.

assumption to be made which greatly simplifies the circuit. In order to avoid induced disturbances in the Kerr-cell circuit the coupling between this part of the circuit and the remainder was made as small as possible. It will therefore be assumed that this coupling is zero. This means that the circuit reduces to that shown in Fig 2. The solution given here will be for the voltage dropping instantaneously. If the solution is desired for any other variation of vol-

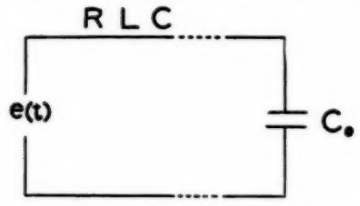


Fig. 2. Kerr cell circuit with spark-gap replaced by $e(t)$.

tage with time it can be obtained by applying the superposition theorem[2] to the solution for instantaneous voltage drop.

In solving this problem the following artifice is used.

$$e_{c_0}(t) = E_0 - e_{c_0}'(t) \tag{1}$$

where $e_{c_0}(t)$ is the voltage on the Kerr-cell, as a function of time, resulting from closing switch in the circuit shown in Fig. 3; E_0 is the initial voltage on

[2] See Operational Circuit Analysis by V. Bush, Page 125.

the Kerr-cell (Fig. 3); and $e_{c_0}'(t)$ is the voltage on the Kerr-cell as a function of time resulting from closing switch in the circuit shown in Fig. 4. It is only necessary therefore to obtain a solution for a voltage applied to the circuit as shown in Fig. 4.

In the solution which follows, the operational method will be used.[2] The plan is briefly: (1) Set up steady state alternating current solution. (2) Replace $j\omega$ in this expression by the operator p. For continuous functions

$$pf(t) = \frac{d}{dt} f(t); \frac{1}{p} f(t) = \int_0^t f(t)dt, \text{ etc.}$$

(3) Carry out the indicated operations according to the rules of operational calculus.

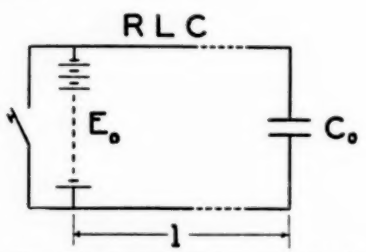


Fig. 3. Schematic Kerr cell circuit representing an instantaneous voltage-drop across spark-gap.

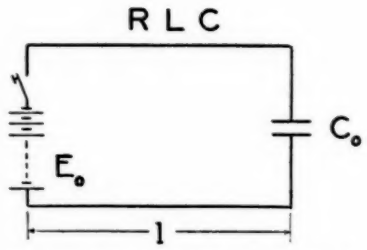


Fig. 4. Schematic Kerr cell circuit representing an instantaneous voltage rise across spark-gap.

Now proceeding in accordance with this plan: the relation between the voltage E_0' and E_{c_0}' for steady state alternating current is given by

$$E_0' = E_{c_0}' \cosh (ZY)^{1/2}l + I_{c_0}' \left(\frac{Z}{Y}\right)^{1/2} \sinh (ZY)^{1/2}l. \tag{2}$$

Where E_0' = applied a.c. voltage

I_0' and E_0' = alternating current and voltage at C_0.

R = resistance per unit length of line

L = inductance per unit length of line

C = capacitance per unit length of line

$\omega/2\pi$ = frequency of E_0'

The current through the condenser may be expressed in terms of the condenser voltage and its impedance:

$$I_{c_0}' = j\omega C_0 E_{c_0}'. \tag{3}$$

Substituting this in Eq. (1) and solving for E_{c_0}

$$E_{c_0}' = \frac{E_0'}{\cosh (ZY)^{1/2}l + j\omega C_0 \left(\frac{Z}{Y}\right)^{1/2} \sinh (ZY)^{1/2}l}. \tag{4}$$

[2] See Operational Circuit Analysis by V. Bush, and Heaviside's Operational Calculus by E. J. Berg.

Replacing $j\omega$ in this expression by the operator p and expressing E_0' as a suddenly applied voltage:

$$e_{c_0}' = \frac{E_0}{\cosh\left((R+Lp)pCl^2\right)^{1/2} + pC_0\left(\frac{R+Lp}{pC}\right)^{1/2}\sinh\left((R+Lp)pCl^2\right)^{1/2}} 1 \quad (5)$$

where 1 represents a function which is zero for all negative values of time and which is unity for all positive values of time.

Rearranging the expressions under the radicals and placing $R/2L=\sigma$ the following expression results.

$$e_{c_0}' = \frac{E_0 1}{\cosh\left(1[(p+\sigma)^2-\sigma^2]CLl^2\right)^{1/2}+\frac{C_0}{C}(LC)^{1/2}((p+\sigma)^2-\sigma^2)^{1/2}\sinh\left([(p+\sigma)^2-\sigma^2]CLl^2\right)^{1/2}} \quad (5')$$

This may be simplified by "shifting" by means of the operational expression

$$\frac{1}{Z(p)}1 = \epsilon^{-bt}\frac{p}{p-b}\,\frac{1}{Z(p-b)}1 \quad (6)$$

or

$$e_{c_0}' = \frac{\epsilon^{-\sigma t}\frac{p}{p-\sigma}E_0 1}{\cosh\left(CLl^2(p^2-\sigma^2)\right)^{1/2}+\frac{C_0}{C}(LC)^{1/2}(p^2-\sigma^2)^{1/2}\sinh\left(CLl^2(p^2-\sigma^2)\right)^{1/2}}.$$

Now the numerator and denominator of this expression will be treated separately and then combined by means of the superposition theorem.[2]

The result of the numerator operating on the unit function is known.

$$\frac{p}{p-\sigma}1 = \epsilon^{\sigma t}. \quad (7)$$

The result of the denominator operating on the unit function may be obtained with the aid of the expansion theorem:

$$\frac{1}{Z(p)}1 = \frac{1}{Z(0)} + \sum_{p=p_1,p_2\cdots}\frac{\epsilon^{pt}}{p\frac{dZ(p)}{dp}} \quad (8)$$

where $Z(0)=Z(p)$ with p placed equal to zero.

p_1, p_2 etc, are the roots of the equation $Z(p)=0$.

$$Z(p) = \cosh\left(CLl^2(p^2-\sigma^2)\right)^{1/2}+\frac{C_0}{C}(LC)^{1/2}(p^2-\sigma^2)^{1/2}\sinh\left(CLl^2(p^2-\sigma^2)\right)^{1/2} \quad (9)$$

$$Z(0) = \cos\left(CLl^2\sigma^2\right)^{1/2} - \frac{C_0}{C}(LC\sigma^2)^{1/2}\sin\left(CLl^2\sigma^2\right)^{1/2}. \quad (10)$$

In order to make the roots of $Z(p)=0$ easier to calculate place $p=jq$.

$$Z(p) = Z(jq) = \cos\left(CLl^2(q^2+\sigma^2)\right)^{1/2} - \frac{C_0}{C}(LC)^{1/2}(q^2+\sigma^2)^{1/2}\sin\left(CLl^2(q^2+\sigma^2)\right)^{1/2} = 0 \quad (11)$$

or

$$\left(CLl^2(q^2+\sigma^2)\right)^{1/2}\tan\left(CLl^2(q^2+\sigma^2)\right)^{1/2} = \frac{Cl}{C_0}. \quad (12)$$

The values of q which are the roots of this equation are to be substituted in the final expression for the voltage e_{c_0}. It will be noted that there are an infinite number of roots and that they are in pairs of plus and minus an imaginary quantity. This means that Eq. (8) may be written as

$$\frac{1}{Z(p)}\,1 = \frac{1}{Z(0)} + \sum_{q=q_1,q_2\cdots} \frac{2\cos qt}{p\dfrac{dZ(p)}{dp}}. \quad (8')$$

The expression $d\,Z(p)/dp$ will now be obtained

$$\frac{dZ(p)}{dp} = \frac{CLl^2\sinh\left(CLl^2(p^2-\sigma^2)\right)^{1/2}}{\left(CLl^2(p^2-\sigma^2)\right)^{1/2}} + \frac{\dfrac{C_0}{C}(LC)^{1/2}(p^2-\sigma^2)^{1/2}(CLl^2p)}{\left(CLl^2(p^2-\sigma^2)\right)^{1/2}}\cosh\left(CLl^2(p^2-\sigma^2)\right) + \frac{\dfrac{C_0}{C}(LC)^{1/2}p}{(p^2-\sigma^2)^{1/2}}\sinh\left(CLl^2(p^2-\sigma^2)\right)^{1/2}. \quad (13)$$

This may be simplified by placing $p=jq$ and then substituting Eq. (11) in Eq. (13). The result is

$$p\frac{dZ(p)}{dp} = -\frac{q^2\left(\dfrac{Cl}{C_0}+1+C_0Ll(q^2+\sigma^2)\right)}{q^2+\sigma^2}\cos\left(CLl^2(q^2+\sigma^2)\right)^{1/2}. \quad (14)$$

Now all the expressions to substitute in Eq. (8′) have been found and it remains to apply the superposition theorem to Eqs. (7) and (8′). The form of superposition theorem most convenient in this case is

$$e(t) = \frac{d}{dt}\int_0^t A_1(t-\lambda)A_2(\lambda)d\lambda \quad (15)$$

where

$$A_1(t) = \frac{1}{Z_1(p)}\,1$$

$$A_2(t) = \frac{1}{Z_2(p)}\,1$$

$$e(t) = \frac{1}{Z_1(p)}\cdot\frac{1}{Z_2(p)}\,1$$

$$e(t) = e_{c_0}' = E_0\epsilon^{-\sigma t}\frac{d}{dt}\int_0^t \epsilon^{\sigma(t-\lambda)}\left[\frac{1}{Z(0)} - \sum\frac{2\cos q\lambda}{p\dfrac{dZ(p)}{dp}}\right]d\lambda. \tag{16}$$

Performing the integration and differentiation this becomes:

$$e_{c_0}' = \frac{E_0}{Z(0)} - E_0\sum\frac{2}{p\dfrac{dZ(p)}{dp}}\,\frac{\sigma^2}{\sigma^2+q^2} - \epsilon^{-\sigma t}E_0\sum\frac{2q\cos\left(qt - \tan^{-1}\dfrac{\sigma}{q}\right)}{(\sigma^2+q^2)^{1/2}p\dfrac{dZ(p)}{dp}}. \tag{17}$$

From physical considerations it is known that for $t=\infty$, $e_{c_0}'=E_0$. The constant term in Eq. (17) is therefore E_0 and Eq. (17) becomes:

$$e_{c_0}' = E_0\left\{1 - \epsilon^{-\sigma t}\sum\frac{2q\cos\left(qt - \tan^{-1}\dfrac{\sigma}{q}\right)}{(\sigma^2+q^2)^{1/2}p\dfrac{dZ(p)}{dp}}\right\}. \tag{18}$$

Now substituting Eq. (14) in Eq. (18) and substituting the result in Eq. (1) the expression desired is obtained.

$$e_{c_0} = \epsilon^{-\sigma t}\sum\frac{2(\sigma^2+q^2)^{1/2}\cos\left(qt - \tan^{-1}\dfrac{\sigma}{q}\right)}{q\left(\dfrac{Cl}{C_0} + 1 + C_0Ll(q^2+\sigma^2)\right)\cos\left(CLl^2(q^2+\sigma^2)\right)^{1/2}} \tag{19}$$

where q is determined from Eq. (12).

It is rather difficult to determine by inspection from this expression just what is the effect of varying the different factors and so a few examples will be given. Before this is done, however, it will be instructive to note the similarity between this expression and the expression derived when the constants are considered as lumped. The solution of the circuit shown in Fig. 5 is

$$e_{c_0} = \frac{E}{(1-\sigma^2LC_0)^{1/2}}\epsilon^{-\sigma t}\cos\left[\left(\frac{1}{LC_0} - \sigma^2\right)^{1/2}t - \tan^{-1}\frac{\sigma}{\left(\dfrac{1}{LC_0} - \sigma^2\right)^{1/2}}\right]. \tag{20}$$

As the ratio of the total distributed capacity of the leads to the capacity of the Kerr cell approaches zero the first term of Eq. (19) approaches Eq. (20) and all other terms approach zero. This means that when the ratio Cl/C_0 is small enough so that $A \tan A = Cl/C_0$ (see Eq. (12)) may be closely enough approximated by $A^2 = Cl/C_0$ the lumped constant solution is a good approximation. The errors involved in this lumped constant solution are; (1) An error in phase, frequency, and magnitude; of the same order of magnitude as the error involved in assuming $A \tan A = A^2$; and (2) An error, due to neglecting higher harmonics, of an amplitude roughly $Cl/5C_0$ times the amplitude of the fundamental.

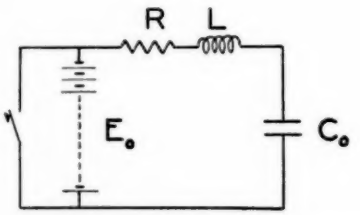


Fig. 5. Schematic Kerr cell circuit with the distributed constants replaced by lumped constants.

To bring out more clearly the characteristics of this circuit and the errors involved in the approximate solution three examples have been selected:

(1) Leads from spark gap to Kerr cell 90 cm long, spaced 15 cm in air, having a total resistance of 150 ohms (both leads).

(2) Leads 90 cm long, spaced 1.3 mm, insulated from each other with mica, a resistance of 150 ohms.

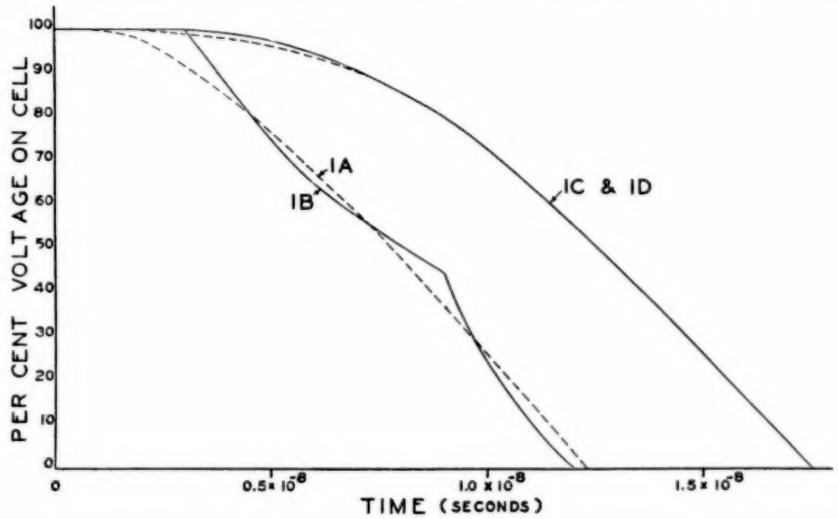
(3) Leads 37 cm long (this length is such that the first wave of voltage arrives at the Kerr cell at the same time that it would if the leads were 90 cm long and had air between them instead of mica), spaced 1.3 mm, insulated from each other with mica, a resistance of 61 ohms (same resistance per cm of leads as in cases (1) and (2).

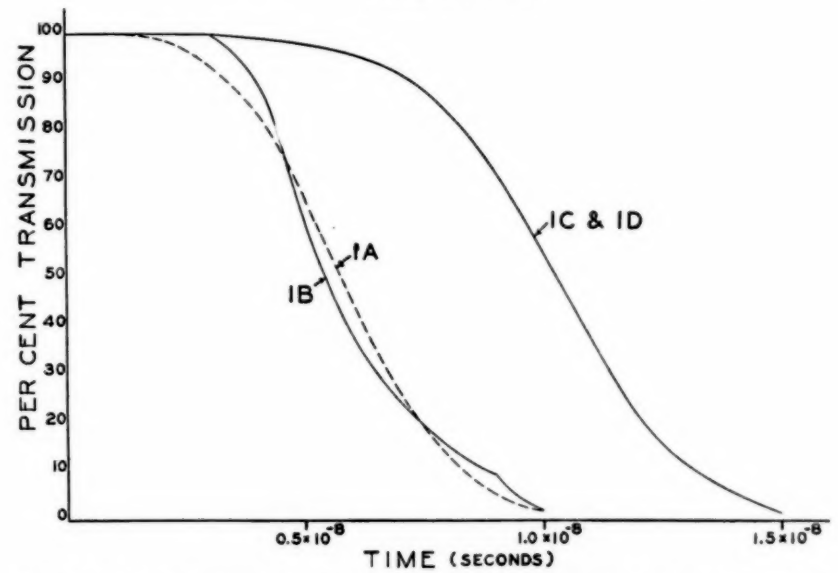
Calculations were made of the voltage on the cell (as a percent, of the break-down voltage of the gap) and the percent transmission[4] of the cell as a function of time for two rates of voltage drop on the spark gap (1). Instantaneous drop of voltage to zero and (2) voltage dropping to zero linearly with time in 10^{-8} seconds.[5] This latter case corresponds roughly to that which we have observed in air with gaps of from 3 to 6 mm. The results of these calculations are given by the curves in Figs. 6 to 10.

[4] The percent, transmission was calculated by the relations $I = \sin^2 \delta/2$, $\delta = 2\pi blE^2$, (where 100 percent transmission was taken as the transmission obtained when $\delta = \pi$ radians, I = transmission, b and l are constants of the Kerr cell, and E = voltage gradient in the cell) and making the assumption that the Kerr cell employed has its maximum opening at the breakdown voltage. If this later condition is not obtained the curves would be somewhat changed but would have in general the same appearance.

[5] The curves for the voltage dropping in 10^{-8} seconds were obtained by applying the superposition theorem to the curves for instantaneous voltage drop. As accuracy was not desired much labor was saved by employing a step by step method which assumed the voltage to fall in five equal instantaneous drops spaced 0.2×10^{-8} seconds apart.

In Figs. 6 and 7 curves 1*a* and 1*b* are for the conditions (1) as stated above and instantaneous voltage drop. Curve 1*a* was computed by the lumped constant equation and 1*b* by the distributed constant equation. For these conditions the difference between these two solutions is sufficiently

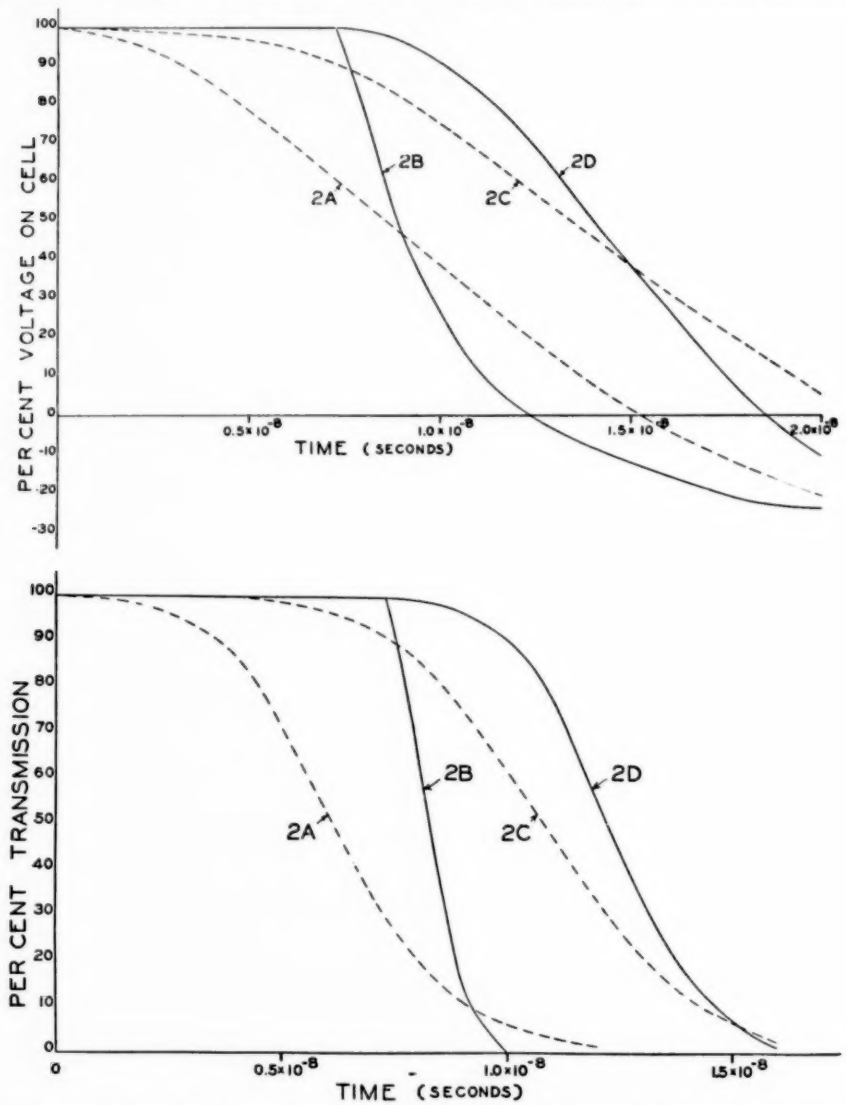




Figs. 6 and 7. Calculated electro-optical shutter performance curves. For 90 cm lead-spaced 15 cm apart in air. 1A, for instantaneous voltage drop across gap—lumped constant solution. 1B, for instantanious voltage drop across gap—distributed constant solution. 1C, for voltage across gap dropping in 10^{-8} seconds—lumped constant solution. 1D, for voltage across gap dropping in 10^{-8} seconds—distributed constant solution.

small to be neglected for our purposes. The breaks in the distributed constant curve at 0.3×10^{-8} and 0.9×10^{-8} seconds are due to the arrival of the first and second waves of current at the Kerr cell.

Curve 1C is for conditions (1) and the voltage dropping in 10^{-8} seconds. The lumped constant and distributed solutions differ so little in this case that they cannot be distinguished on the graph. This is due to the smoothing out of the small irregularities resulting from the less steep wave front associated



Figs. 8 and 9. Calculated electro-optical shutter performance curves. For 90 cm leads spaced 1.3 mm apart on mica. 2A, for instantaneous voltage drop across gap—lumped constant solution. 2B, for instantaneous voltage drop across gap—distributed constant solution. 2C, for voltage across gap dropping in 10^{-8} seconds—lumped constant solution. 2D, for voltage across gap dropping in 10^{-8} seconds—distributed constant solutions.

with the slower rate of voltage drop across the gap. It may be concluded that for leads in air up to one meter in length the lumped constant solution gives a very close approximation of the operation of the circuit.

When the current in the wave front is sufficiently large to make the breaks in the voltage-time curves more pronounced the lumped constant solution no longer gives a good approximation. This is brought out in Figs. 8 and 9 which are for conditions (2). The larger current in the wave front here is due to the closer spacing and to the use of mica both of which increase the distributed capacity. The error which would be involved if the time of operation of the shutter were calculated from the lumped constant equation is shown in Fig. 9 where the percent transmission is plotted vs. time. For instantaneous voltage drop these curves give for time of operation (90 to 10 percent transmission) 5.5×10^{-9} seconds for the lumped and 1.6×10^{-9} seconds for the distributed calulations. For the voltage dropping in 10^{-8} seconds these values are 6.7×10^{-9} seconds for the lumped and 4.6×10^{-9} seconds for the distributed calcu-

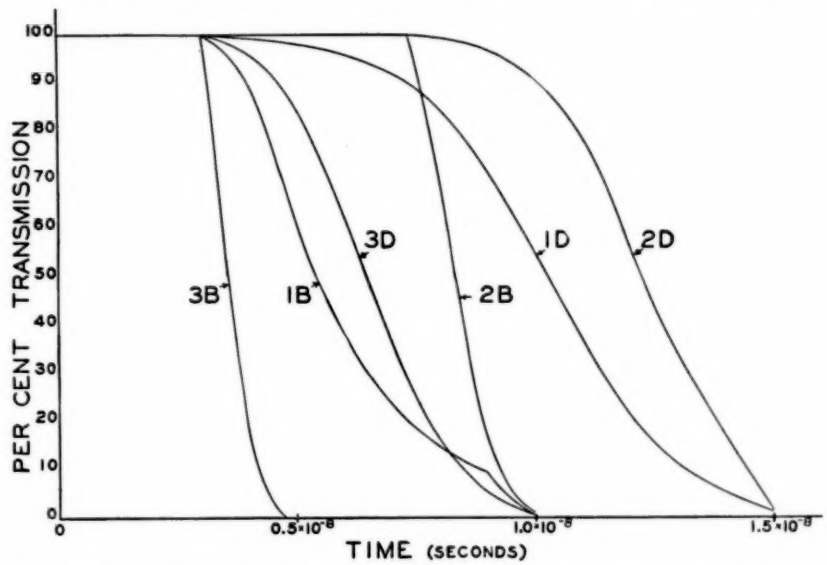


Fig. 10. Calculated electo-optical shutter performance curves. 3B, for 37 cm leads spaced 1.3 mm apart on mica-instantaneous voltage across gap—distributed constants. 3D, for 37 cm leads spaces 1.3 mm apart on mica—voltage across gap dropping in 10^{-8} seconds—distributed constants. 1B, 1D, 2B, and 2D are the same as in Figs. 6, 7, 8, and 9.

lations. It will be seen that the error involved is less for the slower change in voltage but still too large to be neglected. It is interesting to note the extremely short time of operation which is possible even when the voltage drop takes as much as 10^{-8} seconds. That is, if the voltage drops to zero in 10^{-8} seconds the shutter closes in less than half this time. This is due to the fact that the transmission of the cell drops to two percent, when the voltage has dropped to thirty percent, of its initial value.

In order better to compare the operation with the leads in air and the leads on mica, curves were calculated for leads on mica of such a length that the shutter starts to close at the same instant as for leads in air 90 cm long (conditions given under 3 above). This length is given by $90/k^{1/2}$ where k is the dielectric constant of mica. The results of this calculation are given by

Fig. 10. For instantaneous voltage drop on the gap the time of closing of the shutter (90 to 10 percent transmission) for leads in air is 5×10^{-9} seconds and for leads on mica is 1×10^{-9} seconds. If the gap voltage takes 10^{-8} seconds to drop this value is 5.9×10^{-9} for leads in air and 3.8×10^{-9} seconds for the leads on mica. *It is seen that the placing of leads on mica materially shortens the time of closing of the shutter especially if the time of voltage drop across the gap is very small.*

The curves for the 90 cm leads on mica are also drawn in Fig. 10 to bring out the important fact that *the time that the shutter remains open may be increased without appreciably lengthening the time required for the shutter to close.* That is, for the voltage on the gap dropping in 10^{-8} seconds the 37 cm leads give an open time (90 percent transmission or above) of 4.7×10^{-9} seconds and a closing time (90 percent transmission to 10 percent transmission) of 3.8×10^{-9} seconds. The 90 cm leads give an open time of 10.1×10^{-9} seconds and a closing time of 4.3×10^{-9} seconds. The open time is more than doubled while the closing time is increased only 13 percent. This characteristic of the shutter makes possible the observation of various stages of the breakdown with approximately the same sharpness of cut-off.

Experimental Work

It is a simple matter experimentally to measure differences of the open time of the shutter.[6] Referring to Fig. 1., it is seen that two images of the spark will be observed. One image is formed by the light which passes directly from the spark through the Nicol prisms and Kerr cell to the observer. The other is formed by light which first travels to the mirror and is then reflected back through the prisms and Kerr cell to the observer. Different stages of the progression of the spark may be observed in the reflected image by merely moving the mirror nearer to or farther from the spark gap. The difference in time that it takes the spark to progress to these different observed stages is then merely twice the difference of the distance of the mirror from the spark gap divided by the velocity of light. If now it is desired to determine the difference of open time of the shutter for different electrical constants of the Kerr cell spark gap circuit, it is merely necessary to adjust the mirror for each condition of the electrical circuit so that the same state of progression of the spark is observed. The difference of position of the mirror will then give the difference in open time of the shutter for the different conditions of the electrical circuit. This of course assumes that the change in electrical constants does not alter the rate of progression of the spark. The extent to which this is true must be determined experimentally as will presently be shown.

It was desired to determine experimentally the effect of the Kerr cell capacity on the speed of operation of the shutter. To do this a Kerr cell was made which had a constant Kerr effect with a variable capacity. That is, it

[6] The open time of the shutter can also be experimentally determined by a method which makes use of the effect on the brightness of the spark upon receiving the reflected wave from the supply condenser. For an explanation of this method see a paper by Mr. Frank Dunnington in the Phys. Rev. 38, 1506 (1931).

had two fixed plates between which the light passed. In the same cell and connected in parallel (the connecting leads being very short 1 cm) were a fixed and a movable plate by means of which the capacity of the cell was varied.

In order to determine to a fair degree of accuracy the state of progression of the spark the electrodes were chosen so that the spark formed with a narrow bright pencil of light starting from the cathode and moving nearly all the way across the gap before meeting the anode pencil. The state of progression could then be determined by merely measuring with the aid of crosshairs the fraction of the gap which this pencil had travelled. The arrangement of electrodes which gives this condition is a sphere cathode and a plane anode. The cathode used was a sphere 0.5 cm in diameter, the anode a disk 2 cm in diameter, and the spacing was 0.4 cm.

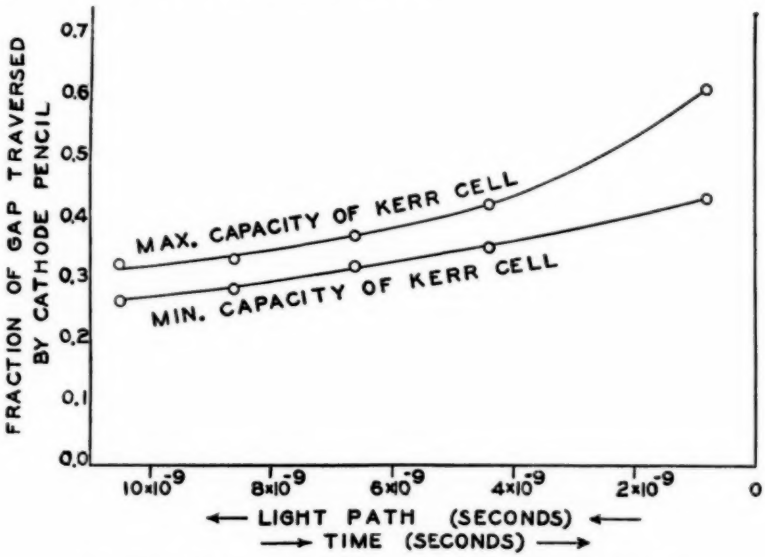


Fig. 11. Experimental curves of progression of spark vs. time.

Three sets of leads were used between the Kerr cell and gap:

(a) 39 cm long 37 ohms resistance (each wire) spaced 15 cm in air.

(b) 39 cm long 84 ohms resistance (each wire) spaced 15 cm in air.

(c) 74 cm long 64 ohms resistance (each wire) spaced 15 cm in air.

Curves of percent, of the gap traversed by the cathode pencil vs. length of light path from gap to Kerr cell (which was varied by means of the mirror) were taken for the minimum and for the maximum capacity of the Kerr cell. The curves thus obtained for the leads *b* are shown in Fig. 11. The abscissa for these curves is plotted from right to left in seconds instead of cm, i.e., the length of light path divided by the velocity of light. Since the longer light paths correspond to earlier times in the formation of the spark the curves from left to right give the progression of the cathode pencil vs. time. If the change in

capacity in the Kerr cell affected the rate of progression of the cathode pencil across the gap the slope of these two curves for a given ordinate would be different. But it is found that if the curves are shifted horizontally they coincide throughout, that is, within the limits of accuracy of the measurements. The change in cell capacity therefore did not measureably affect the rate of formation of the spark in the range for which these curves were taken. Theoretical considerations show us that if the spark were to be effected by the cell capacity it would be effected in the range here shown. These considerations are as follows: The position of the cathode pencil at the instant the shutter closes is given by zero light path or by the zero of the abscissa in Fig. 9. Calculations show that at about 10^{-8} seconds previous to this the voltage across the Kerr cell is 95 percent, of its initial value. Now the only way in which the Kerr cell can affect the spark is by the height of the reflected waves which is in turn determined by the voltage on the cell. Since for times earlier than are shown on the curves this voltage is between 95 and 100 percent, for both minimum and maximum capacity adjustments the spark could not be appreciably affected by a change of cell capacity at these earlier times. This is still more certain when it is realized that the energy fed into the spark by the leads from the supply condenser in these earlier times is about three times as great as that fed in by the Kerr cell circuit.

Since it may now be concluded that the rate of formation of the cathode pencil is the same for both curves in Fig. 9. the difference on the abscissa of the two curves for any given ordinate gives the difference in open time of the shutter resulting from the change in cell capacity.

In order to ascertain whether the Kerr cell and circuit were operating in agreement with the equations developed, the above three differences of time were calculated. The variable Kerr cell had a minimum capacity of 47×10^{-12} farads. This large minimum capacity was made necessary by the variable feature of the cell. For the three sets of leads used in the above measurements the ratio of the total distributed capacity of the leads to the Kerr cell capacity was so small that the lumped constant solution was sufficiently accurate. The next question which was confronted was the time it takes the voltage across the gap to drop. It was known from measurements by other experimenters with the cathode-ray oscillograph that this time was about 10^{-8} seconds.[7] Curves of voltage vs. time were therefore calculated for both instantaneous voltage drop across the gap and for a linear decrease of voltage with time in 10^{-8} seconds. From these curves the difference in open time of the shutter due to the change in cell capacity was determined for: (1) assuming instantaneous voltage drop across the gap, (2) assuming linear voltage drop in 10^{-8} seconds. The result showed that this difference of time for the two cases was the same, or stated better, it changed an amount less than could be distinguished by the experimental measurements. This constancy of the time difference was due to the large capacity of the variable Kerr cell which made the shutter close more slowly than with the regular cell. This situation was very

[7] For a determination of this time with the Kerr cell see article by Mr. Frank Dunnington in Phys. Rev. 38, 1506 (1931).

fortunate as it made the calculated results entirely independant of the way in which the voltage dropped across the gap. This makes possible a mathematical check upon the operation of the Kerr cell and leads alone. Once this agreement is established the calculations may be extended to a case where the rate of voltage drop across the gap has considerable influence and thus determine by trial what the rate of voltage drop is.[7]

The results of the experiments and calculations are given in Table I. It is seen that the results agree within the limits of experimental accuracy. It is unfortunate that the accuracy of the experiment could not have been better.

TABLE I.

Leads	Experimental time difference	Calculated time difference
a	$2.9\times10^{-9}\pm0.5\times10^{-9}$ secs.	2.5×10^{-9} secs.
b	$3.2\times10^{-9}\pm0.5\times10^{-9}$ secs.	2.8×10^{-9} secs.
c	$2.9\times10^{-9}\pm0.5\times10^{-9}$ secs.	3.2×10^{-9} secs.

The percentage error however is not as great as it first appears. The total time open of the cell is about 10^{-8} seconds so that the errors of observation are only about five percent of this value. The remarkable thing is that measurements of such small time intervals in the operation of an optical shutter can even be approximately measured.

At present no quantitative checks have been made on the Kerr cell circuit where the distributed constant solution differs greatly from the lumped constant solution (i.e., when there is a relatively large distributed capacity in the leads). It has however been observed that the cut-off is much sharper when the leads are placed close together and separated by mica. That is, when mica is used the cathode pencil is much more sharply defined at its tip. This is in qualitative agreement with the predictions of the distributed constant solution.

To sum up, it is seen that the accurate solution of the problem of the discharge of the Kerr cell shows that the initial discharge wave causes the cell to close more rapidly than is indicated by the approximate solution involving lumped constants and moreover that the effect of the discharge wave can be enhanced by increasing the distributed capacity along the leads, thereby materially increasing the rate of closing of the shutter.

In conclusion I wish to express my appreciation to Mr. Frank Dunnington for his valuable suggestions and to Professor E. O. Lawrence under whose direction this experimentation is being carried out.

A NEW KIND OF e.m.f. AND OTHER EFFECTS THERMODYNAMICALLY CONNECTED WITH THE FOUR TRANSVERSE EFFECTS

By P. W. Bridgman
Research Laboratory of Physics, Harvard University

(Received January 11, 1932)

Abstract

The Hall e.m.f. attending a current in a magnetic field is subjected to a thermodynamic analysis like that for an ordinary battery, from which it appears that if the Hall e.m.f. has a temperature coefficient, there must be a reversible heating effect when a transverse current flows across a conductor carrying a longitudinal current in a magnetic field. But other arguments show that this heating effect vanishes, and furthermore, it could not be found experimentally. The consequent vanishing of the temperature coefficient of the Hall e.m.f. involves the existence of a new sort of e.m.f., that is, an e.m.f. in a conductor carrying a current in which the temperature is uniformly changing. Corresponding analysis may be made for the other transverse effects. A thermo-motive force connected with the Righi-Leduc coefficient exists in a conductor carrying a thermal conduction current when its temperature changes uniformly. Other relations are deduced connecting the Nernst coefficient and the Ettingshausen coefficient with the new e.m.f. and thermo-motive force. It appears that the temperature dependence of all these quantities is simply connected, and in particular, that the temperature coefficient of the Hall coefficient vanishes at 0°K. The new relations show that certain relations suggested in a previous paper from general considerations of a non-thermo-dynamic character cannot be rigorously exact. A new account is given of the origin of the major part of the Ettingshausen temperature gradient, which is approximately checked by experiment. Finally, the order of magnitude of various small effects is discussed. It is a thermodynamic consequence of the existence of a temperature e.m.f. that there is a temperature change when the current in a conductor changes in magnitude, but it is far below experimental reach. It must be recognized that the specific heat of a conductor is altered by the presence of an electric current. The specific heat is also altered by the presence of an ordinary thermal conduction current. Numerical considerations suggest that the proper velocity to be associated with the thermal current, whether ordinary conduction current, or thermal current convected by an electrical current, is the velocity of sound.

THERE has been so much speculation about the detailed functioning of the mechanisms which may be responsible for the four transverse effects, namely the Hall, Ettingshausen, Nernst, and Righi-Leduc effects, and the whole subject is still in such an unsettled state, that it is well to obtain by arguments of a thermodynamic or other general character all the information which we can which must be independent of any special mechanism. It is not inconceivable that a better understanding of the thermodynamic connections between these effects may lead to a better understanding of the effects themselves. It is surprising how little this method of attack has been used in the past and there are still simple relations of a thermodynamic character which apparently have not been noticed. Practically the only previous applications

of thermodynamics to this subject have been made by Lorentz and myself,[1] but there are still other relations not hitherto touched.

Relations of a purely thermodynamic character may be obtained by constructing electromagnetic or thermodynamic engines utilizing the various effects to furnish energy. Consider first the Hall effect. An electric current flows in a toroid of mean radius a, breadth b, and depth d, breadth and depth being small compared with a. There is a uniform magnetic field of strength H prependicular to the plane of the toroid. The magnetic field may be supposed produced by a permanent magnet with zero temperature coefficient. The circuit is supposed resistanceless, and in the following, irreversible effects arising from the Joulean heating are neglected. This is allowable, because by increasing the linear dimensions of the circuit indefinitely, keeping the total current I constant, the electromagnetic energy of the circuit, $\frac{1}{2}LI^2$, may be made indefinitely large compared with the Joulean dissipation of energy in unit time, RI^2, since L increases in direct proportion to the linear dimensions, and R decreases in the same ratio.

The inner and outer circumferences of the toroid are at a difference of potential in virtue of the Hall effect. If we short circuit across from the inner to the outer circumference, a uniform radial current will flow; this current may be used to drive an external electromagnetic engine, and so energy may be taken out of the system. The external electromagnetic engine may be assumed perfectly efficient, so that the energy output is the product of the Hall potential difference and the amount of electricity flowing transversely. The source of the energy output is primarily the energy of self induction, $\frac{1}{2}LI^2$, associated with the primary current, and the mechanism by which this energy is tapped is the Hall e.m.f. associated with the transverse flow acting circumferentially in the toroid and opposing the primary current. This arrangement is for thermodynamic purposes indistinguishable from a battery, and the ordinary analysis for a battery applies. In particular, if the Hall e.m.f. depends on temperature, then, in analogy with the known behavior of ordinary cells, we may expect reversible heating effects when the transverse current flows. The analysis is so simple that it will pay to reproduce it from the beginning.

Call I the total current, and i the current density, where $I = bdi$. Then the definition of the Hall coefficient at once gives:

$$\text{Transverse e.m.f.} = bHiR,$$

where R is the Hall coefficient, using the conventional notation. If a transverse quantity of electricity dq_e flows, the work done is the product of quantity and e.m.f., or

$$dW = bHiRdq_e.$$

The only variables in this system capable of external manipulation are temperature and transverse flow, which are therefore to be taken as the independent variables. The conservation of energy now gives at once:

[1] H. A. Lorentz, Report of the Fourth Solvay Congress, Conductibilité Électrique des Metaux, 1924, pp. 354–360; P. W. Bridgman, Phys. Rev. **24**, 644–651 (1924).

$$dQ = \left(\frac{\partial E}{\partial \tau}\right)_{q_e} d\tau + \left(\frac{\partial E}{\partial q_e}\right)_{\tau} dq_e + bHiRdq_e,$$

where E is internal energy and Q heat absorbed. Now form $dS = dQ/\tau$, and write down the condition that dS be a perfect differential, by equating the cross derivatives of the coefficients of $d\tau$ and dq_e. This gives at once, neglecting the thermal expansion of the material of the toroid,

$$\left(\frac{\partial Q}{\partial q_e}\right)_{\tau} = \tau bH \frac{\partial}{\partial T}(iR). \tag{1}$$

This indicates that when a transverse current flows in a conductor arranged to show the Hall effect there is a reversible generation of heat required to maintain the system isothermal.

In Eq. (1), Q and q_e are the total amounts of heat and flow of electricity respectively. If Q' is the development of heat per unit volume, we have approximately $Q = 2\pi abdQ'$, and if dq_e' is the density of transverse flow, we have $dq_e = 2\pi dadq_e'$. This gives:

$$\left(\frac{\partial Q'}{\partial q_e'}\right)_{\tau} = \tau H \frac{\partial}{\partial \tau}(iR)_{q_e}, \tag{2}$$

an equation exhibiting the thermal effect in terms of intrinsic properties of the materials, independent of the dimensions of the circuit.

We now have to consider the term $\partial/\partial\tau(iR)_{q_e}$ on the right hand side of the equations. Expanded, this is $i(\partial R/\partial\tau)q_e + R(\partial i/\partial\tau)_{q_e}$. Numerically the proportional change of R for one degree for bismuth, for example, is 0.004. The term $(\partial i/\partial\tau)_{q_e}$ one would probably say on first impulse to be zero, since this denotes the change of current produced by a change of temperature acting so quickly that the Joulean effects may be neglected, and with no transverse flow, that is, with no extraction of work from the system. If the term were not zero, this would demand that there be an e.m.f. in a circuit in which the temperature is changing, and this is an effect not usually considered. I believe, however, that this e.m.f. must exist. My reason is that there are at least two arguments which demand that the absorption of heat accompanying transverse flow be zero, and the only way in which this is reconcilable with the thermodynamic expressions above is that

$$\frac{\partial}{\partial \tau}(iR) \equiv 0, \text{ or } \frac{1}{i}\frac{\partial i}{\partial \tau} = -\frac{1}{R}\frac{\partial R}{\partial \tau} \neq 0.$$

The first argument is derived from the vector character of the current. The transverse and longitudinal currents combine according to the ordinary rules for vectors into a single current, and such a current flowing in a magnetic field is without heating effect, according to original assumption, as far as known experimentally, and also in accordance with the demands of symmetry. The relations here are somewhat simplified by imagining the conductor in the form of a cross, the longitudinal and transverse currents combining

at the center of the cross as indicated to a sufficient degree of approximation in Fig. 1. The only way of saving the situation seems to be to assume that the heating effect is localized in the periphery of the central square, where the direction of current flow changes, as indicated by the dotted lines. But this is inconsistent with the dimensions of the effect as shown by Eq. (2) which exhibits the effect as a heating per unit volume, whereas if the effect were concerned with the change of direction, it would be an effect per unit area.

The second argument is derived from the symmetry of the longitudinal and transverse currents. It is evident in the first place that the heating effect

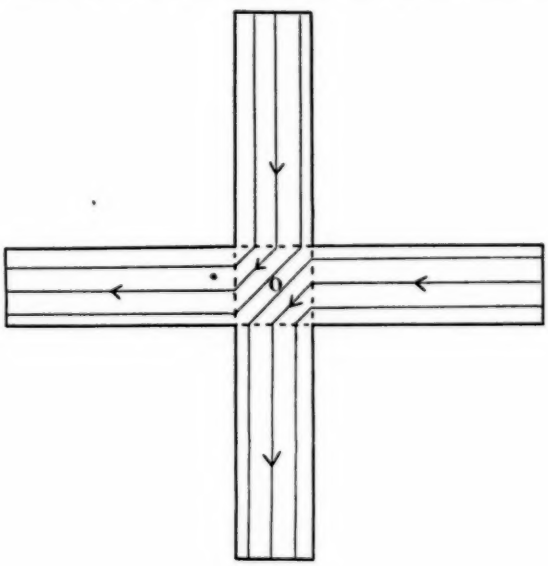


Fig. 1.

of Eq. (2) can be exhibited as a heating effect per unit time per unit transverse current. But if the arrangement is geometrically symmetrical in longitudinal and transverse current, as may be accomplished by making the conductor in the form of a cross as in Fig. 1, then the same result should be obtained independently of which current is called transverse and which longitudinal. An inspection of the figure shows that the symmetry relations make this impossible, for calling the transverse current longitudinal and conversely demands that the effect reverses sign. The only quantity equal to its own negative is zero, showing again that the heating effect must vanish.

There is apparently, therefore, an e.m.f. in a conductor in which the temperature is changing such that

$$\frac{1}{i}\frac{\partial i}{\partial \tau} = -\frac{1}{R}\frac{\partial R}{\partial \tau}.$$

The partial derivative in i may be taken to have the general significance that no work is to be extracted from the system during the change of temperature;

the partial derivative in R may for practical purposes be taken to be the ordinarily determined temperature derivative.

The experimentally determined values of $(1/R)(\partial R/\partial\tau)$ are so large, 0.004 for bismuth, that the related e.m.f. cannot be treated as a negligibly small quantity. An approximate expression for this e.m.f. may be readily found. To get it, we neglect any interaction between the thermal and the electrodynamic energy of the system, setting the total internal energy equal to the sum of the ordinary internal thermal energy in the absence of the current plus the electrodynamic energy, $\frac{1}{2}LI^2$. This amounts to assuming that the specific heat of a conductor carrying a current is the same as that of the same conductor without the current. Later in this paper an estimate will be made of the order of magnitude of this small effect. Utilizing this approximation, the e.m.f. arising from a change of temperature changes only the electrodynamic energy of the system, and we have:

$$\frac{d}{dt}[\tfrac{1}{2}LI^2] = I \times \text{e.m.f.},$$

whence at once:

$$\text{e.m.f.} = L\frac{dI}{dt} = L\frac{\partial I}{\partial\tau}\frac{d\tau}{dt}.$$

But $(1/I)(\partial I/\partial\tau) = (1/i)(\partial i/\partial\tau)$, so that $\partial I/\partial\tau = -I(1/R)(\partial R/\partial\tau)$, and

$$\text{e.m.f.} = -\left(LI\frac{1}{R}\frac{\partial R}{\partial\tau}\right)\frac{\partial\tau}{dt}. \tag{3}$$

That is, the temperature e.m.f. in a circuit in which the temperature is changing at unit rate is $-LI(1/R)(\partial R/\partial\tau)$.

One can see in a general way why there should be an effect of this kind. In the first place, it arises from the magnetic field of the current on itself, the external magnetic field having dropped out of the picture. That the external field ought to have no net effect is suggested by the theorem of elementary electrodynamic theory that there is no mutual energy between an electrical current and a system of permanent magnets. Some of the electrons which constitute the current will move perpendicular to the magnetic field of the current itself, and will thus experience an action in virtue of the Hall effect which will have a component along the original current. The intensity of this action involves the self magnetic field, which explains how L gets into the picture. Furthermore, the number and distribution of the transversely moving electrons is a function of the temperature, so that there will be an interaction, manifesting itself as an e.m.f., when temperature changes. To give a detailed account of this effect from the statistical point of view would probably be prohibitively complicated, and would involve integration over the entire conductor of many terms, a number of which would drop out from the final result.

An experimental attempt was made to detect the existence of the transserve heating effect, when I first noticed the analogy with an ordinary bat-

tery, and had not yet realized that the relations between other quantities were such as to make this zero. Two crosses of bismuth like Fig. 1 were cast, the arms being about 1.2 by 0.6 cm in section. In casting, they were chilled rapidly from the molten condition so as to make the crystal structure as fine grained as possible. These crosses were mounted face to face, separated by a layer of cellophane for insulation, and the two junctions of a copper-constantan thermocouple were attached to the two centers of the crosses, indicated by 0 in the figure. The couple indicated, therefore, the differential effect at the centers of the two crosses. The electrical connections were such that the currents in the various branches could be varied independently, both as to magnitude and direction. The magnetic field was about 5000 gauss, and the maximum current density about 10 amperes per square cm. The difficulty with the experiment is in eliminating the effect of the finite size of the crystal grains. Because of the unequal resistance of the grains in different directions, the current experiences many internal changes of direction. Each of these is accompanied by a local heating effect, in virtue of the internal Peltier heat. This is changed by the application of a magnetic field, because of the effect of the field on the resistance. Effects of this kind exist with only a longitudinal current. However, by using all possible combinations of currents, and noting that some of the effects change sign when current direction changes and some do not, it was possible to show that if any heating effect of the kind corresponding to Eq. (2) exists it must be less than 10 percent of that part of the non-isotropic effects which is due to the action of the magnetic field. Numerically, this meant that any final shift of equilibrium of temperature due to the effect sought was less than 0.006°. This was much less than a preliminary calculation had indicated was to be expected on the assumption that $\partial i/\partial \tau = 0$. This is probably as good a proof of the non-existence of the effect as can be given without very much more elaborate precautions. It was a great surprise to find that the magnetic influence on the internal non-isotropic effects was so large. It raises the question whether such effects have been sufficiently considered in previous measurements; measurements of the Ettingshausen temperature difference would be particularly susceptible to error from this source.

Returning now to the relation $(\partial/\partial\tau)(iR) = 0$, we can derive a suggestion as to the behavior of R at 0° K. It seems highly probable that the e.m.f. arising from change of temperature vanishes at 0°K. We would expect this from general considerations suggested by experience with the third law, and the mechanistic explanation of this e.m.f. just given would suggest the same thing. For the transverse components of motion of the electrons which constitute the current would be expected to lose all their haphazard quality at low temperatures, and therefore their capacity for taking part in thermal effects. If $\partial i/\partial\tau$ vanishes at 0°K, this demands:

$$\lim_{\tau=0} \frac{1}{R} \frac{\partial R}{\partial \tau} = 0.$$

This relation appears to be consistent with the experimental results found

at Leiden,[2] although the experimental accuracy is not always great enough to give perfectly definite indications.

The other transverse effects may now be subjected to an analysis similar to that above for the Hall effect. In doing this it will be convenient to introduce new coefficients in place of the conventional Nernst and Righi-Leduc coefficients, since the conventional definitions of these involve a lack of symmetry as compared with the Hall and Ettingshausen coefficients. The conventional Nernst coefficient, Q_N is to be replaced by Q_N' where $Q_N' = Q_N/k$, k being the thermal conductivity, and the conventional Righi-Leduc coefficient S_R is to be replaced by S_R', where $S_R' = S_R/k$. These coefficients are written with subscripts N and R to avoid confusion with the thermodynamic symbols Q for quantity of heat and S for entropy. These altered definitions now give the following consistent scheme for the four transverse effects:

(1) Hall transverse potential gradient with longitudinal electric current, $i, = RHi$.

(2) Ettingshausen transverse temperature gradient with longitudinal electric current, $i, = PHi$.

(3) Nernst transverse potential gradient with longitudinal heat current, $w, = Q_N'Hw$.

(4) Righi-Leduc transverse temperature gradient with longitudinal heat current, $w, = S_R'Hw$.

i and w are here density of electrical and thermal current.

Imagine now the toroid of the preceding analysis with a circumferential heat current replacing the electrical current. The ring will have to be split along some radius and the two sides of the slit maintained at a difference of temperature. This temperature difference is to be maintained irrespective of how the mean temperature of the whole system may change. The irreversible effects connected with thermal conduction in such a system may be neglected by making all changes in the system rapidly, so that the dissipation due to the thermal conduction is vanishingly small compared with other effects.

The exact parallel of the preceeding analysis for the Hall effect may now be made. Allow a quantity of heat dq_w to flow transversely, and utilize this to drive a thermodynamic engine working between the temperature limits of the Righi-Leduc temperature difference. This difference is $S_R'bHw$, and the work received from the engine is:

$$dW = \frac{1}{\tau} S_R'bHwdq_w.$$

The source of this work is the work done by the longitudinal heat current in flowing through the longitudinal Righi-Leduc temperature difference accompanying the flow of the transverse heat current. A result exactly similar to that before follows at once on writing down the condition that the entropy change be a perfect differential, namely:

[2] Bengt Beckman, Leiden Communications, Supplement, No. 40, 1915.

$$\left(\frac{\partial Q}{\partial q_w}\right)_\tau = \tau b H \frac{\partial}{\partial \tau}\left(\frac{w S_R'}{\tau}\right),$$

or this may be written as heat per unit volume in terms of densities:

$$\left(\frac{\partial Q'}{\partial q_w'}\right)_\tau = \tau H \frac{\partial}{\partial \tau}\left(\frac{w S_R'}{\tau}\right). \tag{4}$$

Exactly the same arguments as applied before, namely one from combining longitudinal and transverse heat currents vectorially, and one from the effect of interchanging longitudinal and transverse currents, may be applied to this case, showing that this heating effect must vanish, giving the relation $\partial/\partial\tau(wS_R'/\tau)=0$, or:

$$\frac{1}{\frac{S_R'}{\tau}} \frac{\partial}{\partial \tau}\left(\frac{S_R'}{\tau}\right) = -\frac{1}{w}\frac{\partial w}{\partial \tau}. \tag{5}$$

It is to be presumed that S_R'/τ varies with temperature, and that therefore the term $(1/w)(\partial w/\partial\tau)$ exists. This is the formal analogue of the expression $(1/i)(\partial i/\partial\tau)$, and denotes a change in a thermal current when the mean temperature is changed, no external work being taken from the heat current and the change being made so rapidly that the dissipation of the thermal stream against thermal resistance is negligible. Such phenomena connected with thermal currents certainly have not been detected, and the mechanism must be quite different from that in the electrical case. An electrical current is capable of coasting for a certain time after the e.m.f. has ceased, driven by the stored energy of self induction. The strict analogue of self induction does not exist for a thermal current. If, however, the thermal current is at all like ordinary currents in having a property analogous to velocity, it must also have a space density, so that the energy content and therefore the specific heat of a body carrying a thermal current is different from that of an equivalent assembly of infinitesimal elements with no thermal current. This space density of energy may perform the same function as the energy of self induction of an electrical current, and give meaning to the derivative $\partial w/\partial\tau$. An estimate will be made later of the order of magnitude of such effects. They are too small to be detected by direct experiment, but we may nevertheless recognize their existence and use them in theoretical discussion.

The other two effects, the Ettingshausen and Nernst effects may be similarly analyzed. By allowing a transverse heat current to flow in the presence of a longitudinal electrical current, energy may be taken out of the system in virtue of the Ettingshausen transverse temperature difference, and by allowing electricity to flow transversely in the presence of a longitudinal heat current energy may be taken out in virtue of the Nernst transverse potential difference. The same analysis as before demands accompanying heat effects, which, written for unit volume, are respectively:

$$\left(\frac{\partial Q'}{\partial q_w'}\right)_\tau = \tau H \frac{\partial}{\partial \tau}\left(\frac{iP}{\tau}\right) \tag{6}$$

$$\left(\frac{\partial Q'}{\partial q_e'}\right)_\tau = \tau H \frac{\partial}{\partial \tau}(wQ_N'). \tag{7}$$

We expect as before that these two effects are zero in virtue of other relations. The argument has to be somewhat modified, however. The first argument disappears entirely, because a heat current and an electric current do not combine vectorially. The second argument may, however, be appropriately modified. It is to be noticed in the first place that there is a reciprocal relation between the sources of the energy of the phenomena involved in Eqs. (6) and (7). The energy extracted by the transverse thermal flow of Eq. (6) is provided by the longitudinal electric current flowing against the Nernst e.m.f. acting longitudinally associated with the transverse heat current. Similarly, the energy extracted by the transverse electric flow of Eq. (7) has its source in the longitudinal heat current flowing against the Ettingshausen temperature difference acting longitudinally associated with the transverse electric flow. Eq. (6) now demands that dQ' be positive when the transverse heat current extracts energy from the longitudinal electric current, and (7) demands a positive dQ' when the transverse electric current extracts energy from the longitudinal heat current. But the situation of Eq. (6) may also be described as a longitudinal heat current in the presence of a transverse electric current, and the energy relations demand that the transverse electric current give energy to the longitudinal heat current. We thus again have the dilemma of a quantity equal to its own negative, and the only way out is the quantity itself to vanish. Hence

$$\frac{\partial}{\partial \tau}\left(\frac{iP}{\tau}\right) = 0 \tag{8}$$

and

$$\frac{\partial}{\partial \tau}(wQ_N') = 0. \tag{9}$$

The $\partial i/\partial \tau$ which occurs in (8) is the same as that which occurred in connection with the Hall coefficient, and the $\partial w/\partial \tau$ of (9) is the same as in the expression for the Righi-Leduc coefficient. Eliminating these derivatives gives:

$$\frac{1}{\frac{P}{\tau}} \frac{\partial}{\partial \tau}\left(\frac{P}{\tau}\right) = \frac{1}{R} \frac{\partial R}{\partial \tau},$$

and

$$\frac{1}{\frac{S_R'}{\tau}} \frac{\partial}{\partial \tau}\left(\frac{S_R'}{\tau}\right) = \frac{1}{Q_N'} \frac{\partial Q_N'}{\partial \tau}.$$

Integration gives at once:

$$\frac{P}{\tau} = \text{const}_1 R,$$

and

$$\frac{S_R'}{\tau} = \text{const}_2 Q_N'.$$

The constants are independent of temperature, but may of course vary from substance to substance.

Furthermore, $Q_N' = P/\tau$, as was shown in the preceding paper to be demanded by the energy relations. We therefore have the relations:

$$R = \text{const}_3 \frac{P}{\tau} = \text{const}_3 Q_N' = \text{const}_4 \frac{S_R'}{\tau}. \tag{10}$$

That is, R, P/τ, Q_N', and S_R'/τ all depend on temperature in the same way, and therefore, in particular, all vanish in the same way at 0°K.

In addition to the relation $P/\tau = Q_N'$ deduced in the previous paper from the first law of thermodynamics, two other relations were also deduced from much more doubtful premises, such, for example, as the assumption that the rotation of the equipotential lines is the fundamental feature of the Hall effect, and is the same whether the potential drop is an iR drop as in an ordinary conductor, or whether it somes from a Thomson effect in an unequally heated bar. These two other relations were: $Q_N' = \sigma/k\rho R$, and $P = \sigma\tau S_R'$, where σ is the Thomson coefficient and ρ specific electrical resistance. Consistency of these relations with those above would demand that $\sigma/k\rho = \text{const}$, and $\sigma = \text{const}/\tau$. But $1/k\rho$ is the Wiedemann-Franz ratio, and varies approximately inversely as τ, so that the first relation demands that σ vary directly as τ, while the second demands that it be inversely proportional to τ. One or both of the previous relations must be given up. It is probable that neither is exactly correct, because the experimental evidence would not seem to indicate that σ varies either directly or inversely as τ in general. Of the two relations, the first, $Q_N' = (\sigma/k\rho)R$, rests on the more questionable argument and seems definitely not to agree with experiment in the case of metallic Co. At the time of writing that paper the relation was apparently satisfied for Co, using the value of Moreau[3] for the Nernst coefficient, Q_N'. Professor E. H. Hall, however, was of the opinion that the sign of Moreau's Q_N' was incorrect and checked this opinion by a redetermination of Q_N' experimentally. The only way of saving the relation was therefore by assuming that the usual sign of σ for Co is incorrect. I made a redetermination of σ on two pieces of Co cut longitudinally and transversely from the same specimen as used by Professor Hall, and verified that σ has the accepted sign. These measurements have not been previously reported. According to the best experimental evidence, therefore, the relation definitely fails for Co, and of course cannot be regarded as a general relation. It is still noteworthy, however, that the rela-

[3] G. Moreau, discussed on page 227 of the book by L. L. Campbell, Galvanomagnetic and Thermomagnetic Effects, Longmans, 1923. The numerical values and notation of this paper are taken from Campbell's book.

tion is satisfied by most other metals within experimental error, which must be admitted to be large.

At present I can see no argument of a plausible character suggesting another relation to replace $Q_N' = (\sigma/k\rho)R$. The second relation, however, $P = \sigma\tau S_R'$, may be replaced by another very similar to it with much plausibility. In a paper on conclusions to be drawn from the existence of various thermo-electric phenomena in crystals[4] I showed that the electric current I, must be recognized to convect with it a thermal current of magnitude $I\tau\int_0^\tau \sigma d\tau/\tau$. In particular, the longitudinal electric current of the Hall effect convects with it this amount of thermal energy. This heat flow will give rise to a transverse temperature gradient by the Righi-Leduc effect. If we make the simple assumption that this is the entire transverse temperature gradient when the longitudinal electric current flows, we have at once a connection between the Ettingshausen and the Righi-Leduc coefficients, namely:

$$P = \tau S_R' \int_0^\tau \frac{\sigma d\tau}{\tau},$$

which differs from the relation previously proposed only in that σ is replaced by $\int_0^\tau \sigma d\tau/\tau$. One would expect these two quantities to be of approximately the same magnitude; examination of the experimental evidence will show that the experimental accuracy is not great enough to give much significance to apparent differences, and that the new relation may be considered to be verified by experiment with the same degree of accuracy as the old relation. Consistency of the new relation with Eq. (10) demands that $\tau\int_0^\tau \sigma d\tau/\tau$ be constant, which gives on integration, $\sigma = \text{const}/\tau$. This cannot be a rigorously correct relation, because, among other things, it would give an infinite value to the integral at the lower limit, and therefore an infinite thermal energy convected by the electric current. The Ettingshausen temperature gradient cannot, therefore, all originate in the simple way suggested, but a large part of it must be of this origin, as shown by the approximate experimental check of the relation between P and S_R'.

Finally, we try to form an idea of the magnitude of the small effects neglected in the argument above. There is in the first place a heating effect associated with a change in the current. Imagine a closed electric circuit carrying a current i, in which the current is maintained by the self induction L, and in which the Joulean dissipation may be neglected. This system is determined thermodynamically by its temperature τ and the current i, and these may be taken as the independent variables fixing the state of the system. Temperature may be varied by any conventional means; i may be varied by inserting into the circuit an e.m.f. ϵ through which the system delivers or receives work from the outside, with accompanying change of i. The rate at which work is exchanged with the surroundings is, under these conditions:

$$\frac{dw}{dt} = \epsilon i.$$

[4] P. W. Bridgman, Phys. Rev. **31**, 221–235 (1928).

The change of i during the action of ϵ is governed by the ordinary equation of balance of e.m.f.'s, which in this case gives the statement that ϵ is balanced by the e.m.f. of self induction and the temperature e.m.f. already discussed. This gives:

$$L\frac{di}{dt} + iL\frac{1}{R}\frac{dR}{d\tau}\frac{d\tau}{dt} = \epsilon.$$

The first law of thermodynamics gives:

$$dQ = dW + dE,$$

where E is the internal energy, or,

$$dQ = \epsilon i dt + dE.$$

Put $dE = \partial E/\partial\tau d\tau + \partial E/\partial i di$, and substitute the value for ϵ, obtaining,

$$dQ = \left(\frac{\partial E}{\partial \tau} + i^2 L\frac{1}{R}\frac{dR}{\partial \tau}\right)d\tau + \left(\frac{\partial E}{\partial i} + iL\right)di.$$

Form dS by dividing by τ, and formulate the condition that this be a perfect differential as usual by equating the cross derivatives. This gives at once:

$$\left(\frac{\partial Q}{\partial i}\right)_\tau = -2\tau iL\frac{1}{R}\frac{dR}{d\tau}.$$

This gives the inflow of heat required to maintain the system isothermal when the current is altered. By dividing by the heat capacity we can at once obtain the approximate change of temperature when the current is altered adiabatically. For ordinary circuits such temperature effects are very small. Thus, in a toroid of bismuth of 10 cm radius and 1 cm thickness, the change of temperature when current is increased from 0 to 100 amp/cm^2 is of the order of 10^{-13} degrees centigrade.

The story is not completely told by these considerations. To characterize completely the system it would be necessary to determine the internal energy as a function of temperature and current. It may be shown by considerations which need not be given in detail here that the internal energy is not merely additive of the ordinary electrodynamic energy $\frac{1}{2}Li^2$ and the ordinary thermal energy when there is no current, but there must be cross terms. This means that the specific heat depends on the current. The precise influence of the current on the specific heat is not determined by the phenomena hitherto discussed, but apparently involves a new constant of the substance, which must be determined by independent experiment. We may attempt an estimate of the order of magnitude as follows. We have already seen that a current convects with it a thermal energy $\tau\int_0^\tau \sigma d\tau/\tau$; a good enough approximation for our purpose to this somewhat complicated expression is $\tau\sigma$. In the metal cobalt σ is unusually large, being 2.2×10^{-6} volts/°C. This means that one coulomb convects with it the thermal energy $(273\times2.2\times10^{-6})/4.2 = 1.4\times10^{-4}$

gm cal., which means that when the current density is 1 amp./cm^2, 1.4×10^{-4} gm cal. of thermal energy is convected across each square centimeter in one second. This energy flow is to be thought of as having a velocity and therefore a space density, and it is in virtue of the space density that the specific heat of a conductor carrying a current is different from that of the conductor without the current. Both the velocity and the corresponding space density are entirely unknown. Perhaps the most immediate assumption is that the velocity is the same as that of the electrons which constitute the current. This may be computed by assuming, as in the Sommerfeld theory, that the number of conduction electrons is the same as the number of atoms. This gives a velocity of 6×10^{-5} cm/sec for Co, which means a density of energy of $1.4\times10^{-4}/6\times10^{-5}=2$ gm cal./cm^3. So large a value appears inadmissible, since it would demand measureable effects on the specific heat, and would also demand that the apparent self induction of a circuit be appreciably dependent on the temperature as well as on the geometrical dimensions. If, on the other hand, the proper velocity to be associated with the convected thermal energy is the velocity of sound propagation, as it is for an ordinary thermal current, the effects will be smaller by a factor of 10^{11}, and therefore beyond experimental reach.

A corresponding analysis may be carried through for thermal flow, but the corresponding effects are more difficult to visualize because of the absence of a thermal self induction. If however, the space density of energy associated with the thermal current is taken as the analogue of the energy of self induction, corresponding results may be found. It will be found that there is a "thermo-motive" force in a body carrying a thermal current when the temperature changes. Analysis like that for the electrical case gives a simpler result because the energy associated with the thermal current is to be taken as proportional to the thermal current, rather than proportional to the square, as in the electrical case. This will lead to a heating effect in a substance in which the thermal conduction current is altered, which turns out to be proportional to the absolute temperature. But all this is so far beyond the reach of experiment that it is of little profit to pursue the matter further. It is, however, perhaps of interest to attempt to form an idea of the order of magnitude of the space density of energy associated with a thermal current. Imagine a centimeter cube of copper between the opposite faces of which there is a temperature difference of 100°. The thermal flux is approximately 100 cal./sec. The volume density of energy corresponding to this flux is such that its product into the velocity of flux is equal to 100. For the velocity we may take, in accordance with the Debye picture of thermal conduction, the velocity of sound, which for copper is about 3.5×10^6 cm/sec. The space density of energy is therefore $100/3.5\times10^6=3\times10^{-5}$ gm cal./cm^3. The heat capacity of 1 cm^3 of copper is about 0.8 gm cal. This means, therefore, that if a copper cube in which a thermal current of 100 cal./sec is flowing is suddenly isolated from the source and sink of heat flow, its final equilibrium temperature will be about 4×10^{-5} °C higher than its average temperature during the flow.

This, of course, would be very difficult to detect. It is interesting, however, that if a different velocity were assumed, as for example a velocity of the order of a few cm per sec., which is the order of the apparent velocity with which the maxima or minima of ordinary periodic thermal disturbances sink into the metal, a temperature effect of the order of many degrees would have been found. This affords rather direct confirmation of the correctness of the Debye point of view. The experiment might be worth making to find how far the velocity limit could be pushed.

ELECTRICITY AS A NATURAL PROPERTY OF RIEMANNIAN GEOMETRY

By Cornelius Lanczos
Department of Mathematics, Purdue University

(Received October 21, 1931)

Abstract

Field equations of a Riemannian geometry which can be deduced from a Hamiltonian principle have the following general property: Due to the conservation law of Einstein's curvature tensor R_{ik} there appears in the integration of the field equations a free vectorial function ϕ_i, determined, however, by the conservation law. This has been shown by applying a mathematical formulation of Mach's principle to the extremely weak deformation of a given arbitrary metric. Specifying the Hamiltonian function by the evident condition of gauge invariance this free vectorial function ϕ_i has all the fundamental properties of the electro-magnetic vector potential: the law of continuity is strictly fulfilled everywhere, the potential equation is to be deduced in first approximation and also the Lorentz' ponderomotive force of a particle. In this theory the material particle is to be considered as a proper solution of the field equations.

SEVERAL attempts have been made since Einstein's discovery of the theory of general relativity to find a solution of the problem "electricity" in a way analogous to that employed in the successful solution of the problem gravitation. These attempts are mostly characterized by the idea of enlarging the geometrical basis of Riemannian geometry with the tendency to find in this way a geometrical formation corresponding to the vector potential or to the antisymmetrical tensor of electromagnetic field-strength. The following development is characterized by the fact that Riemannian geometry is retained without any modification, discovering in this geometry itself an analytical element, which seems to correspond with a surprising harmony to the electromagnetic vector potential.

1. The Field Equations $R_{ik}=0$ as Characterization of Euclidian Geometry.

The tendency of Einstein has been to generalize the equations $g_{ik}=\text{const}$ by setting up a system of differential equations of second order, usually written in the form:

$$R_{ik} = 0. \tag{1.1}$$

The gravitational field of the sun, the moving of planets around the sun may be perfectly described by these equations.[1] However, we have to remark that such a solution of the field equations, is not without singularities. From the

[1] Even the assumption, the path of a planet may be a geodesic is not a new hypothesis, the dynamics of a mass-point being a consequence of the field equations. Cf. Einstein, Berl. Akad. Ber. 1927, p. 2; Lanczos, Zeits. f. Physik 59, 514 (1930).

point of view of these equations matter appears as a singularity. Although from a purely mathematical standpoint the use of singularities is not to be rejected, several features suggest the uselessness of singularities in the description of nature. The singularity is always to be considered as the *failing* of a law. For example, to illustrate the situation by a simple familiar example: it makes no difference whether we say that we have the potential equation

$$\Delta\phi = 0 \tag{1.2}$$

permitting singularities or that we have Poisson's equation

$$\Delta\phi = -4\pi\rho \tag{1.3}$$

without singularities. The latter form of the equation shows that in some regions the law $\Delta\phi = 0$ is not fulfilled, since the function ρ in some regions differs from 0. In the same way Einstein's law (1.1) gives a Riemannian space but only if we admit singularities, in other words exceptions form the postulated law, since the singularity is equivalent to the fact that in some regions the tensor R_{ik} differs from 0. If we exclude singularities and demand the fulfilling of the law $R_{ik} = 0$, in the *whole space*, adding some natural boundary conditions practically equivalent with the fact that even in infinity the law (1.1) may be fulfilled, we obtain *Euclidian geometry*. And so Einstein's curvature tensor R_{ik} is to be considered as a *fundamental formation* of Riemannian geometry, being the simplest invariant characterization of a geometry. We are accustomed to consider the Riemannian curvature tensor $R_{\alpha\beta\gamma\delta}$ as a perfect characterization of geometry, the vanishing of this tensor being a proof for Euclidian geometry, while the vanishing of R_{ik} is possible also in a curved world. But the consideration of the Riemannian tensor is only necessary considering certain *limited* regions of space. In this case the vanishing of R_{ik} may not be sufficient to introduce Euclidian geometry. But this insufficiency disappears considering the *whole* space without any interruption. The characterization of a geometry by the Riemannian tensor is indeed not of a natural kind, because it shows a high degree of *over-determination*. It is not satisfying to characterize the metric tensor g_{ik} which is a symmetric tensor of second degree by a tensor of fourth degree. Einstein's curvature tensor R_{ik}, however, has just the right degree of determination being also a symmetric tensor of second degree exactly corresponding to the fundamental metric tensor.

2. General Form of Field Equations for R_{ik} which have a Hamiltonian Principle

We shall consider Einstein's curvature tensor R_{ik} as a fundamental concept of Riemannian geometry, able to characterize such a geometry. The simplest statement $R_{ik} = 0$ gives only the Euclidian geometry, for this reason we expect field equations for the R_{ik} corresponding to the general character of Riemannian geometry which is a differential geometry. We will suppose that these field equations are to be deduced from a Hamiltonian principle. The tensorial problem is then reduced to the consideration of a single in-

variant, the Hamiltonian function. This Hamiltonian function will be a function of the R_{ik}, but being an invariant it must contain also the g_{ik}. Any non-covariant components may then be deduced to the pure covariants using the g_{ik} writing for example

$$R^{ik} = R_{\alpha\beta}g^{\alpha i}g^{\beta k} \tag{2.1}$$

thus we can write our Hamiltonian function in the following form:

$$H = H(R_{ik}, g_{ik}). \tag{2.2}$$

Our action integral is:

$$I = \int H dv \tag{2.3}$$

dv being the volume element and the condition of minimum is the familiar:

$$\delta I = 0 \tag{2.4}$$

corresponding to an arbitrary variation of g_{ik}. To find our field equations in a useful form we shall consider in the process of variation at first the R_{ik} and the g_{ik} as *independent* variables. Doing that we can write the variation of our integral generally in the following form:

$$\delta I = \epsilon \int (\mu^{ik}\rho_{ik} - v^{ik}\gamma_{ik})dv \tag{2.5}$$

where ϵ is an infinitesimal parameter, the variation of R_{ik} is denoted by ρ_{ik}, the variation of g_{ik} with γ_{ik}. We do not specify at first the Hamiltonian function and use therefore the general expressions u_{ik} and v_{ik}, which are easy to find if the Hamiltonian function is given.

The variation of R_{ik} and the variation of g_{ik} are related in a covariant way. The relation is given in a former investigation of the author, exploring the extremely weak fields in Einstein's theory.[2] We use the resulting formulae of this work without proof, the details of calculation may be found in the quoted paper.

We use the following notations. The operation $\partial/\partial x_i$ shall mean always a *covariant differentiation*, not the common differentiation. We introduce the invariant Laplace operator:

$$\Delta = g^{\alpha\beta}\frac{\partial^2}{\partial x_\alpha \partial x_\beta} \tag{2.6}$$

and

$$E(\gamma_{ik}) = \Delta\gamma_{ik} + R_i{}^\alpha\gamma_{k\alpha} + R_k{}^\alpha\gamma_{i\alpha} - 2R_{i\alpha k\beta}\gamma^{\alpha\beta} \tag{2.7}$$

$$D(\gamma_{ik}) = E(\gamma_{ik}) - \left(\frac{\partial\chi_i}{\partial x_k} + \frac{\partial\chi_k}{\partial x_i}\right) \tag{2.8}$$

[2] Lanczos, Zeits. f. Physik **31**, 112 (1925).

where the vector χ_i has the significance

$$\chi_i = \frac{\partial \gamma_i{}^{\alpha}}{\partial x_\alpha} - \frac{1}{2}\frac{\partial \gamma}{\partial x_i} \qquad (\gamma = \gamma_\alpha{}^{\alpha}). \tag{2.9}$$

We need also the "adjoint" expression of $D(\gamma_{ik})$:

$$F(\gamma_{ik}) = E(\gamma_{ik}) - \left(\frac{\partial \sigma_i}{\partial x_k} + \frac{\partial \sigma_k}{\partial x_i} - \frac{\partial \sigma^{\alpha}}{\partial x_\alpha} g_{ik}\right) \tag{2.10}$$

with

$$\sigma_i = \frac{\partial \gamma_i{}^{\alpha}}{\partial x_\alpha}. \tag{2.11}$$

The relation between ρ_{ik} and γ_{ik} is expressed by[3]

$$\rho_{ik} = \tfrac{1}{2} D(\gamma_{ik}). \tag{2.12}$$

Corresponding to the general properties of the "adjoint" differential expressions we can write

$$\int [u^{ik} D(\gamma_{ik}) - \gamma^{ik} F(u_{ik})]\, dv = \text{surface int.} \tag{2.13}$$

According to this equation we can transform the term with ρ_{ik} in our expression (2.5) and set equal 0 the resulting coefficient of γ_{ik}. We obtain in this way the field equations in the following form:

$$F(u_{ik}) = 2v_{ik} \tag{2.14}$$

3. A General Method of Integration. Appearance of a Free Vectorial Function ϕ_i in the Integrated Form of Field Equations

We can use also another method to find our field equations. If there is a unique connection between ρ_{ik} and γ_{ik} we can also choose the ρ_{ik} as the primary variables reducing the γ_{ik} to the ρ_{ik}. For the vanishing of δI this change cannot make any difference this vanishing being required for *any* variation. The physical meaning of this process corresponds to the idea that Einstein called the "principle of Mach".[4] This principle attempts to determine the metric of a manifold from a given distribution of matter characterized by the material tensor. This principle has a precise significance only in the case of a *weak* deformation of a given metric field and the variation corresponds to just such a supposition. Mathematically we have to solve the differential equation (2.12) determining the γ_{ik} from the ρ_{ik} by integration.

This process is indeed possible with a certain natural restriction. An arbitrary deformation γ_{ik} can be of two kinds: a real and an apparent deformation. The latter results from pure transformations of coordinates and has no importance. Its form is:[5]

[3] Reference 2, Eq. (16).
[4] Einstein, Ann. d. Physik **55**, 241 (1918).
[5] Reference 2, Eq. (30).

$$\gamma_{ik} = \frac{\partial \Phi_i}{\partial x_k} + \frac{\partial \Phi_k}{\partial x_i} \tag{3.1}$$

where Φ_i is an arbitrary vector. It is evident that we should exclude such apparent deformations, which neither produce matter nor influence our invariant integral.

We do that by *normalizing* our coordinate system. We can always add to a given γ-field an apparent field of the character (3.1) by making a suitable transformation of coordinates. We can now determine the vector Φ_i so that the vector χ_i, defined by (2.9) becomes 0 for the resulting field.

$$\chi_i = \frac{\partial \gamma_i{}^{\alpha}}{\partial x_\alpha} - \frac{1}{2}\frac{\partial \gamma}{\partial x_i} = 0 \tag{3.2}$$

that gives a vectorial differential equation for the Φ_i which is to be solved. The only condition is that the homogeneous equation does not have "proper solutions", i.e., any regular solution besides 0. This homogeneous equation is:

$$\frac{\partial \gamma_i{}^{\alpha}}{\partial x_\alpha} - \frac{1}{2}\frac{\partial \gamma}{\partial x_i} = 0 \tag{3.3}$$

with the assumption (3.1) for γ_{ik}. That gives[6]

$$\Delta \Phi_i - R_i{}^{\alpha} \Phi_\alpha = 0. \tag{3.4}$$

In order to apply Mach's principle we require that this equation does not have any solution besides $\Phi_i = 0$.

After normalising our coordinates we have between ρ_{ik} and γ_{ik} the simple connection

$$E(\gamma_{ik}) = 2\rho_{ik} \tag{3.5}$$

which differential equation is now *self-adjoint*. We can integrate this equation uniquely, and thus find a unique correspondence between γ_{ik} and ρ_{ik}, if the homogeneous equation

$$E(\gamma_{ik}) = 0 \tag{3.6}$$

has no proper solutions besides 0. However, a special proper solution cannot be avoided:

$$\gamma_{ik} = \kappa g_{ik} \tag{3.7}$$

where κ is a constant—disturbing the general claim of our method. But the significance of this exception is very easy to discover. The interpretation of (3.7) is a change in the *gauge*. That is not a real deformation and therefore does not produce any matter. On the other hand, it does not belong to the pure transformations of coordinates, because the value of ds^2 is changed, owing to the fact that the measure of a length is dimensioned and so depends upon the gauge used. Since this exceptional deformation is trivial and disappears if we normalize our gauge, its existence is not to be considered as an

[6] Reference 2, Eq. (32).

objection against Mach's principle. We see that our program to use Mach's principle becomes possible by normalizing the coordinates and normalizing the gauge. While the first normalization can be produced in a natural way by annuling the vector χ_i, the latter normalization remains artificial. But the change in the gauge is certainly without any influence on the result if we have a Hamiltonian integral which is "gauge-invariant", that means it does not depend on the gauge. We will use later just this principle to determine our Hamiltonian function. And it is therefore sufficient for our purpose to use any *arbitrary* normalization.

We have now attained the desired unique correspondence between ρ_{ik} and γ_{ik} and will only state that the possibility of accomplishing our program is given through the conditions that the equation (3.4) does not have any solution besides 0 and the equation (3.6) does not have any solution besides (3.7) with a constant κ. We will add the remark that a proper solution imposes always a condition for the right side of a differential equation: the condition of orthogonality. The homogeneous equation (3.6) having the proper solution (3.7) we get the following scalar condition which must be fulfilled by the material tensor ρ_{ik}:

$$\int \rho dv = 0 \quad (\rho = \rho_\alpha{}^\alpha). \tag{3.8}$$

The solution of the differential equation (3.5) by integration will be symbolized thus:

$$\gamma_{ik} = T(\rho_{ik}) \tag{3.9}$$

where T is a certain integral operator, the inverse operator of E.

Using this equation we are ready to consider the ρ_{ik} as the primary quantities, considering γ_{ik} as a function of the ρ_{ik}. However, we must notice an important fact. The γ_{ik} have been treated as completely arbitrary quantities. We made later a restriction excluding the apparent deformations, but this restriction does not influence the result. Regarding the ρ_{ik}, however, the situation is quite different. We cannot consider the material tensor as completely arbitrary, it must always obey the general vectorial conservation law of momentum and energy. This law is a mathematical identity in Einstein's theory of relativity and expresses a fundamental property of Riemannian geometry. As it is necessarily fullfilled also for the deformed geometry, we must observe it even during the variation. The condition which must be fulfilled in every point of the manifold is given by the following equation:[7]

$$\frac{\partial \rho_i{}^\alpha}{\partial x_\alpha} - \frac{1}{2}\frac{\partial \rho}{\partial x_i} = P_{\alpha\beta i}\gamma^{\alpha\beta} \tag{3.10}$$

putting

$$P_{\alpha\beta i} = \frac{1}{2}\left(\frac{\partial R_{i\alpha}}{\partial x_\beta} + \frac{\partial R_{i\beta}}{\partial x_\alpha} - \frac{\partial R_{\alpha\beta}}{\partial x_i}\right). \tag{3.11}$$

[7] Reference 2, Eq. (25). It is supposed that we use the normalized system of coordinates $\chi_i = 0$.

We have to consider this equation as an *auxiliary condition* of our variation problem, corresponding to the ingenious general method of Lagrange: the method of undetermined multipliers. Since in our case the condition is vectorial, the multiplier will be a *vector*, we denote it with $2\phi_i$. According to Lagrange's method we add to our variation integral the expression

$$2\int\left[\phi^i\left(\frac{\partial \rho_i{}^\alpha}{\partial x_\alpha}-\frac{1}{2}\frac{\partial \rho}{\partial x_i}\right)-P_{\alpha\beta i}\gamma^{\alpha\beta}\phi^i\right]dv. \tag{3.12}$$

In the same way we must treat the condition (3.8) which also imposes a restriction on the ρ_{ik}. This latter condition introduces a constant λ as Lagrangian factor and we have to add the expression

$$\lambda\int \rho_{ik}g^{ik}dv \tag{3.13}$$

to our variation. We are now authorized to treat our problem completely like a free variation problem without any restrictions, and we can equate according to the usual method the coefficients of ρ_{ik} to 0. But, since the derivatives of ρ_{ik} appear in the first term of (3.12) we will make a partial integration in the familiar way, finding,

$$-\int \rho^{ik}\left(\frac{\partial \phi_i}{\partial x_k}+\frac{\partial \phi_k}{\partial x_i}-\frac{\partial \phi^\alpha}{\partial x_\alpha}g_{ik}\right)dv. \tag{3.14}$$

Thus we find our field equations in the following form:[8]

$$U_{ik}=T(V_{ik}) \tag{3.15}$$

if we put

$$U_{ik}=u_{ik}+\lambda g_{ik}-\left(\frac{\partial \phi_i}{\partial x_k}+\frac{\partial \phi_k}{\partial x_i}-\frac{\partial \phi^\alpha}{\partial x_\alpha}g_{ik}\right) \tag{3.16}$$

$$V_{ik}=v_{ik}+2P_{ik\alpha}\phi^\alpha. \tag{3.17}$$

Finally we can transform the obtained Eq. (3.15) in a pure differential equation. We invert the integral operator T in the corresponding differential operator E, observing the two Eqs. (3.5) and (3.9) which are equivalent. We obtain then our field equations in the following form:

$$\left.\begin{aligned} U_{ik}&=2w_{ik}\\ E(w_{ik})&=V_{ik}\end{aligned}\right\}. \tag{3.18}$$

The Eqs. (3.15) and (3.18) are equivalent. Compared with the original form (2.14) they represent a new formulation of the field equations. We have in fact but a new *form*, not a new *statement*, using in both cases the same Hamiltonian principle, only changing the variables, which process cannot influence the result. The difference is that the new formulation is an *integrated* form of our original field equations (2.14). This integration would certainly

[8] We employ in the deduction the property of T to be a *symmetric* operator: $\bar{T}=T$.

be possible also by a direct investigation of the field equations in its original form. But the derivation in an indirect manner, using the variation principle itself by means of Mach's principle, is much shorter, giving the result immediately.

The striking and important result is, that the integrated form of the field-equations shows the *appearance of a free vectorial function* ϕ_i. The necessity to introduce it lies in the conservation law which demands that we regard it as an auxiliary condition for the variation. We have to expect, therefore, that the same condition must give us a determination for the undetermined Lagrangian factor ϕ_i. That is indeed the fact. We have to consider our field-equations first of all as determining equations for the fundamental curvature tensor R_{ik}. To find a metric belonging to this tensor is only possible if the conservation law is fulfilled. It is not to be expected that an arbitrary solution of our field equations will possess this quality but the possibility of attaining this must exist. And that can occur by the use of the free function ϕ_i which has to be determined in such a manner that the conservation law becomes fulfilled. This gives just the right determination for ϕ_i, namely a vectorial differential equation in every point of our manifold. Anticipating the later physical interpretation we will call the vectorial function ϕ_i appearing as a free function in the integration of field equations and determined by the conservation law: the "vector potential".

It suggests itself to raise the question: What may be the reason that the original field equations of gravitation $R_{ik}=0$ do not show the appearing of any vector potential although according to the general theorem the vector potential appears always upon integrating our equations. The only condition for this theorem, viz., the existence of Hamiltonian principle, is fulfilled in the case of the equations $R_{ik}=0$.

The reason of this fact is very peculiar. The general method leads in this case to the following connection between curvature tensor and vector potential:[9]

$$R_{ik} = \frac{\partial \phi_i}{\partial x_k} + \frac{\partial \phi_k}{\partial x_i} . \tag{3.19}$$

This connection corresponds perfectly to the connection (3.1) between an apparent deformation γ_{ik} produced by a pure transformation of coordinates, and the arbitrary vector ϕ_i. The conservation law, the divergence equation of the tensor $R_{ik}-\frac{1}{2}Rg_{ik}$ corresponds completely to the homogeneous Eq. (3.3) upon substituting the expression (3.2). One of the conditions for carrying out Mach's principle was that this equation may not have any solution besides 0. As a consequence this follows

[9] Lanczos, Zeits. f. Physik **32,** 163 (1925); Eq. (33). This apparent *enlargement* of Einstein's equations baffled the author the first time when he discovered the above explained method in the just mentioned paper. He thought that using Mach's principle we are able to *complete* the field equations by the vector potential. This erroneous idea was the hindrance to discover the general importance and real meaning of this method, namely to be a general method of *integration* without any change in the field equations.

$$\phi_i = 0. \tag{3.20}$$

And so just in the case of the gravitational equations the vector potential *remains latent*, disappearing owing to the conditional equation which it has to fulfill.

4. Gauge-Invariance of the Hamiltonian Intergral

The fact that the Riemannian ds^2 is a *dimensioned* quantity and therefore furnished with an arbitrary factor cannot be avoided. It is, however, a natural assumption that our action integral should not be affected by this arbitrariness. We have to seek a minimum for a certain function viz., the action integral. If there is an arbitrary gauge-factor influencing the value of this function, the minimum problem loses its meaning because it is possible to obtain any value by a proper choice of gauge. It is natural to require that the action-integral should not depend on the gauge, i.e. it has to be a *non-dimensioned* quantity, a pure number. In this case the arbitrariness of gauge is overcome; ("gauge-invariance"). In 4 dimensions the unit of volume element dv is $[L]$.[4] On the other hand the unit of R_{ik} is $[L]^{-2}$; (irrespective of the number of dimensions of the space). We see that the dependency of the Hamiltonian function on R_{ik} in 4 dimensions has to be of a *quadratic* kind.

We find only *two* invariants of the required quality built up upon the R_{ik}:[10]

$$H_1 = R_{\alpha\beta}R^{\alpha\beta} \tag{4.1}$$

$$H_2 = (R_{\alpha\beta}g^{\alpha\beta})^2 = R^2. \tag{4.2}$$

Therefore we should choose a linear combination of these two invariants.

$$H = H_1 + CH_2 \tag{4.3}$$

as our Hamiltonian function. C is a numerical constant whose value is not to be decided at present. It seems that the quality of the two possible functions is connected with the duality of electricity and gravitation in nature. The first by itself leading, as we will show, to the familiar phenomena of electricity, the second by itself leading to the pure gravitational equations. The combination of both seems to be necessary to build up a material particle as a "proper solution" of the field equations.

We are now able to replace in the field equations the previously undetermined u_{ik} and v_{ik} by the specified values which we obtain from the specified Hamiltonian function. We only need to carry out the variation with respect to R_{ik} and g_{ik}, considering these two kinds of quantities as independent of each other. We find without any difficulty the following expressions. Using the first invariant:

$$u_{ik} = 2R_{ik} \tag{4.4}$$

$$v_{ik} = 2\left(R_{i\alpha}R_k{}^{\alpha} - \tfrac{1}{4}R_{\alpha\beta}R^{\alpha\beta}g_{ik}\right) \tag{4.5}$$

[10] The invariant $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ is likewise gauge-invariant but does not correspond to our point of view, because the Riemannian curvature tensor $R_{\alpha\beta\gamma\delta}$ cannot be reduced to Einstein's curvature tensor R_{ik}.

Using the second invariant:

$$u_{ik} = 2Rg_{ik} \tag{4.6}$$

$$v_{ik} = 2R(R_{ik} - \tfrac{1}{4}Rg_{ik}). \tag{4.7}$$

The construction of v_{ik} produced by the first invariant shows a great analogy to Maxwell's stress-tensor being built up in the same way upon the symmetric tensor R_{ik} as Maxwell's tensor upon the antisymmetric tensor F_{ik} the electromagnetic field-strength. This analogy becomes even greater considering the fact, which we shall deduce later, that to the first approximation we have

$$R_{ik} = \frac{\partial \phi_i}{\partial x_k} + \frac{\partial \phi_k}{\partial x_i}$$

while on the other hand F_{ik} is the same combination but antisymmetrically:

$$F_{ik} = \frac{\partial \phi_i}{\partial x_k} - \frac{\partial \phi_k}{\partial x_i}.$$

We observe the remarkable quality that the scalar $v = v_\alpha{}^\alpha$ *vanishes* in the first as well as in the second case.

$$v = v_\alpha{}^\alpha = 0. \tag{4.8}$$

5. First Integral of our Field-Equations

The last mentioned quality of the v_{ik} permits us to find a scalar integral of our field equations. We consider the field equations in their original form (2.14). We build the scalar equation

$$g^{ik}F(u_{ik}) = 2g^{ik}v_{ik}. \tag{5.1}$$

The right side vanishes in consequence of (4.8). The left side is reduced to a single term if we utilize the fundamental conservation law:

$$\frac{\partial R_i{}^\alpha}{\partial x_\alpha} - \frac{1}{2}\frac{\partial R}{\partial x_i} = 0 \tag{5.2}$$

and it yields

$$\Delta R = 0. \tag{5.3}$$

This equation has undoubtedly the solution

$$R = \text{const.} \tag{5.4}$$

But that is also the *only* solution. Make the assumption given in (3.7) for the homogeneous Eq. (3.6). That leads to the condition

$$\Delta \kappa = 0. \tag{5.5}$$

Any solution of this equation besides $\kappa = \text{const.}$ would produce a possible solution of (3.6) which we have excluded beforehand. The solution (5.4) of (5.3) is therefore not only possible but also *necessary*. In this way we have found a first integral of our field equations of a scalar kind: the *scalar Riemannian curvature must be constant in every point of our manifold.*

6. A Further First Integral: the Equation of Continuity for the Vector Potential

We now use the *integrated* form of our field equations, the equations (3.18), building in the same way the corresponding scalar equation on multiplying by g^{ik}. The second equation gives

$$w = w_\alpha{}^\alpha = \text{const.} \tag{6.1}$$

Introducing that in the first equation we find

$$2R(1 + 4C) + 4\lambda + 2\frac{\partial \phi^\alpha}{\partial x_\alpha} = \text{const.} \tag{6.2}$$

We are surprised that we do not come back to our former result $R = \text{const.}$ But the difference is that in the earlier deduction we used not only the field equations but also the divergence condition of R_{ik} which is not utilized for the deduction of (6.2). But we can introduce now our former result (5.4) and attain in this way

$$\frac{\partial \phi^\alpha}{\partial x_\alpha} = \text{const.} \tag{6.3}$$

This constant cannot differ from 0, otherwise the solution $\phi_i = 0$ would not be possible and the magnitude of the vector potential would increase more and more. The only possibility is therefore

$$\frac{\partial \phi^\alpha}{\partial x_\alpha} = 0. \tag{6.4}$$

That is the well-known divergence condition of the vector potential corresponding to the conservation law of electricity in the electromagnetic field theory.

7. The Potential Equation in First Approximation

The constant of integration in (5.4) must be extremely small. Otherwise the average curvature-radius of the universe could not be so extremely great. For this reason the second invariant recedes to the background in significance compared with the first because the u_{ik} and v_{ik} produced by it, contain this very small value as a factor. To suppose that the constant C could compensate this small value and make the second invariant dominant by a choice of a very high value for it seems to be very improbable. In this case the gravitational force would have to dominate the electrical force but in reality the reverse is true. The importance of the second invariant may be connected with the existence of proper solutions[11] which are to be considered as the representation of a material particle. But beyond the very central parts of matter with which there is possibly connected a very high curvature its influence seems to be negligible, or at most to be regarded as a correction. The essen-

[11] That means non-trivial solutions of the field equations which are everywhere regular.

tial part of the Hamiltonian function is H_1, and we will content ourselves with considering only this part.

The field equations are to be written:[12]

$$\left.\begin{aligned} R_{ik} &= \frac{1}{2}\left(\frac{\partial \phi_i}{\partial x_k} + \frac{\partial \phi_k}{\partial x_i}\right) + w_{ik} \\ E(w_{ik}) &= 2(R_{i\alpha}R_k{}^{\alpha} - \tfrac{1}{4}R_{\alpha\beta}R^{\alpha\beta}g_{ik} + P_{ik\alpha}\phi^{\alpha}) \end{aligned}\right\}. \tag{7.1}$$

We observe that w_{ik} is a quantity of *second* order if we consider R_{ik} to be of the first order. We can therefore write as a first approximation:

$$R_{ik} = \frac{1}{2}\left(\frac{\partial \phi_i}{\partial x_k} + \frac{\partial \phi_k}{\partial x_i}\right) \tag{7.2}$$

this connection between curvature tensor and vector potential corresponds to the connection (3.19), deduced for the case of Einstein's gravitational equations as mentioned above. However, in our case this connection is only a *first approximation* not a strict law. The correctious terms of second degree which have to be added become high in the central field and remove the necessity that the vector potential has to vanish.

On constructing the divergence equation we find in first approximation the following determining equation for ϕ_i:

$$\Delta \phi_i = 0 \tag{7.3}$$

which is the familiar potential equation. The notation "vector potential" is therefore indeed justified.

8. The Ponderomotive Lorentz Force in First Approximation

A formation like the antisymmetric tensor of the electromagnetic field-intensity

$$F_{ik} = \frac{\partial \phi_i}{\partial x_k} - \frac{\partial \phi_k}{\partial x_i} \tag{8.1}$$

nowhere appears in our treatment. The question arises then how can we give an explanation for this quantity. If we find a possibility of deducing a dynamical law of motion for a material particle like the phenomenological Lorentz law, we will be indeed justified to identify the vectorial function, ϕ_i appearing in the integration of our field equations, with the electromagnetic vector potential. Because all the fundamental facts of the electromagnetic field are embraced by the potential equation and the law of continuity for the vector potential, and also the Lorentzian force expressing the action of the field upon the material particle. The first two are already deduced. Finally, we consider the last problem, viz. the action of the field on a particle.

[12] We neglect also the "cosmological term" λg_{ik} which is only important for large regions of the universe. It is interesting to remark that the field equations themselves do not allow any artificial assumption to introduce the cosmological constant, it appears as *constant of integration*.

A special dynamical law independent on the field equations does not exist in a theory built up on the idea of proper solutions excluding any singularities. The field equations are able to determine every event and we do not have the arbitrariness connected with the behavior of singularities. The influence of an external field on the motion of a particle has to be considered as a consequence of the field equations.

We cannot, however, proceed directly in this way to find a dynamical law. We do not at present have sufficient mathematical tools for the treatment of such a complicated system of equations especially those concerning the problem of proper solutions. On the other hand: we do not need a *detailed* knowledge about the influence of an external field on the field of a proper solution. Our desire is only to obtain the resulting influence on the particle *as a whole*. For this purpose it seems much more convenient to avoid the direct use of the field equations and to attain the goal by a treatment which has also proved to be the best way to integrate the field equations without directly using them. This treatment is the direct application of Hamilton's principle. We know that this principle is valid for *every* variation. We are thus justified in applying it to the special variation which sufficies to obtain the dynamical law of a material particle. We do not need for this purpose the detailed knowledge of the construction of a proper solution which represents the material particle. Some general features are sufficient to find an approximate solution of our problem. That is the great advantage of this process.

The second invariant H_2 is without importance for our purpose. In consequence of the intermediary integral $R=\text{const.}$ it becomes a constant. We can then observe this condition also during the variation so that the variation of this part of the action-integral disappears and we have to consider only the first integral.

We consider our field as a superposition of the proper field of the particle and a weak external field. The proper field produces an integral which can approximately be calculated on supposing that the perturbance of the external field is only weak and the reaction of the particle on itself is negligible ("quasi-static" condition). We know that the central high part of the field is chiefly important in its contribution to the integral and this central field is to be considered as moving with the particle unchanged and not essentially modified by the weak external field. Supposing that the average distribution of the field may have a spherical symmetry we can write the resulting integral in the following form:

$$m \int ds \tag{8.2}$$

with the constant m on which turns out later to be the "electro-magnetic mass"; ds is the line-element of the world-line of the particle.

We have to calculate now the interaction between the internal and the external field. We cannot survey the situation in the high central field. But it seems probable to consider this part of interaction as to be weak as compared with the interaction in the "ether", meaning with this word the very large

exterior part of space in which the field becomes weak, and where the linear approximation is permitted and the principle of superposition is valid. The predominance of this part of space with respect to the interaction is a consequence of its huge extension compared with the "matter", this word meaning the high central field of small extension in which no superposition is possible. The duality "field-matter" is in this sense retained although obviously without precise distinction.

In this "ether" we are able to accomplish our calculation. We have here corresponding to (7.2)

$$R_{\alpha\beta}R^{\alpha\beta} = \frac{1}{4}\left(\frac{\partial\phi_\alpha}{\partial x_\beta} + \frac{\partial\phi_\beta}{\partial x_\alpha}\right)^2 \tag{8.3}$$

the metric being in first approximation Euclidian. The interaction is given by

$$\begin{aligned} I_{12} &= \frac{1}{2}\int\left(\frac{\phi_\alpha{}^{(1)}}{\partial x_\beta} + \frac{\partial\phi_\beta{}^{(1)}}{\partial x_\alpha}\right)\left(\frac{\partial\phi_\alpha{}^{(2)}}{\partial x_\beta} + \frac{\partial\phi_\beta{}^{(2)}}{\partial x_\alpha}\right)dv \\ &= \int\left(\frac{\partial\phi_\alpha{}^{(1)}}{\partial x} + \frac{\partial\phi_\beta{}^{(1)}}{\partial x_\alpha}\right)\frac{\partial\phi_\alpha{}^{(2)}}{\partial x_\beta}\,dv. \end{aligned} \tag{8.4}$$

The index 1 means the internal, the index 2 the external field. With aid of the potential equation and the continuity law for $\phi_i{}^{(2)}$ we can transform this volume integral into a surface integral to be taken over the boundary surface of the considered region:

$$I_{12} = \int\phi_\alpha{}^{(2)}\nu_\beta\left(\frac{\partial\phi_\alpha{}^{(1)}}{\partial x_\beta} + \frac{\partial\phi_\beta{}^{(1)}}{\partial x_\alpha}\right)dF \tag{8.5}$$

dF is the surface-element, (the "surface" is here, of course, a 3-dimensional region) ν_i the normal of the surface at the considered point. The boundary surface is a small tube around the material particle separating the ether from the matter. The positive direction of ν_i shows towards the inside of the tube.

We can accomplish our integral in two steps, integrating at first around the small cross-section of the tube, and secondly, over the world-line of the particle. During the first step we can consider $\phi_i{}^{(2)}$ in this small region of the space as a constant putting it in front of the integral sign. We have only to consider the first integral, the second vanishing, owing to the law of continuity which is fulfilled also on the inside of the tube. This first integral can be written in the following form

$$\int\phi_\alpha{}^{(2)}\sigma_\alpha ds \tag{8.6}$$

where σ_i is a vector, depending only on the particle itself. If we suppose that the field of the particle has a spherical symmetry, the direction of σ_i can only be the direction of the world-line. And going in a "rest-system" we see that the length of σ_i is nothing else but the electric *charge* of the material particle:

$$\bar{\sigma}_4 = \int \frac{\partial \bar{\phi}_4^{(1)}}{\partial x_\alpha} \nu_\alpha df = e. \tag{8.7}$$

The dash should mark that we are in a rest system. df is the surface-element of a common surface (of our 3-dimensional space) enclosing the electron.

We therefore have to put

$$\sigma_i = e\dot{x}_i. \tag{8.8}$$

(The point over a letter means: differentiation with respect to ds) and our whole action integral becomes

$$m \int ds + e \int \phi_\alpha^{(2)} \dot{x}_\alpha ds. \tag{8.9}$$

Using the familiar variation-equations of Euler-Lagrange we obtain immediately

$$m\ddot{x}_i + e\left(\frac{\partial \phi_i^{(2)}}{\partial x_\alpha} - \frac{\partial \phi_\alpha^{(2)}}{\partial x_i}\right)\dot{x}_\alpha = 0 \tag{8.10}$$

which is nothing else but the *ponderomotive law of Lorentz*. Here, for the first time, appears the antisymmetric combination (8.1) which never appeared earlier and has no fundamental significance.

It seems that the vector ϕ_i, originally a free vector of integration, arising and also determined by the fact that a fundamental quantity of Riemannian geometry, viz., the curvature tensor of Einstein obey the law of conservation, corresponds completely to the fundamental quantity of the electromagnetic field viz., the vector potential. We get the impression that it is not only a mistake to believe that the Riemannian geometry contains nothing like electricity, but on the contrary the existence of electricity, seems to be a direct test for a fundamental property of Riemannian geometry the conservation law of Einstein's curvature tensor R_{ik}.

It should be necessary to investigate whether there are proper solutions of the field-equations with the charges $-e$ and $+e$ but different masses, representing the fundamental material particles electron and proton.

9. Some Additional Remarks Concerning the Mathematical Foundation of the Theory[13]

If we consider the homogeneous equation

$$F(u_{ik}) = 0 \tag{9.1}$$

[13] This chapter, written in the beginning of December, contains some further improvements of the author in the subject presented above. The formulation of this chapter is influenced by some critical remarks of Professor H. P. Robertson, who examined the contents of the previous chapters with great care. He deduced also independently of the author the fundamental Eq. (9.11) for the vector potential, recognizing at the same time an error in the sign of a formula in a previous paper of the author. (See the next footnote). The last chapter 10, which establishes the connection with Maxwell's equations, was added January 1932.

of the field Eqs. (2.14) and we suppose a Euclidian metric we find a solution of the form

$$u_{ik} = \frac{\partial \phi_i}{\partial x_k} + \frac{\partial \phi_k}{\partial x_i} - \frac{\partial \phi^\alpha}{\partial x_\alpha} g_{ik} \tag{9.2}$$

with an arbitrary vector ϕ_i. For this reason we should expect that the quite similar appearance of the free vector ϕ_i in the integrated form (3.18) of the field equations (2.14) should be connected with the fact that the homogeneous equation has a solution with an arbitrary vector ϕ_i. That was originally my conjecture. As a consequence of that it would be necessary that there exist a vectorial identity for $F(u_{ik})$ and corresponding to that a vectorial condition for the right side of the equation. However, we cannot prove the existence of such an identity, and also, substituting our solution (3.18) for the special case $v_{ik} = 0$ in our field equations we do not find that the homogeneous Eq. (9.1) is necessarily satisfied. The real situation is the following.

We introduced in our procedure the idea that a pure transformation of the coordinates cannot influence our variation principle which is an invariant. We used this fact to restrict the variation of g_{ik} by the condition of (3.2) which cannot change anything in our results. However, while this treatment is justified if our integral of variation is deduced by a scalar Hamiltonian function $H(R_{ik}, g_{ik})$, we do not have the same statement in the "non-holonomic" case, i.e. if u_{ik} and v_{ik} are considered as *general* quantities. In this case the exclusion of the pure transformation of coordinates is a *real* restriction and treating this restriction as an auxiliary condition in the variation of g_{ik} we obtain our field equations instead of (2.14) in the more general form

$$F(u_{ik}) - 2v_{ik} = \frac{\partial \Phi_i}{\partial x_k} + \frac{\partial \Phi_k}{\partial x_i} - \frac{\partial \Phi^\alpha}{\partial x_\alpha} g_{ik} \tag{9.3}$$

with the free vector Φ_i as a Lagrangian factor. It therefore follows that already in the *original* form of the field equations there appears a free function Φ_i, if we operate with the general quantities u_{ik} and v_{ik}. Assuming the existence of a Hamiltonian function, the vector Φ_i disappears automatically. We know namely in this case that the left side of (9.3) satisfies an identity, viz.,

$$\operatorname{div} [F(u_{ik}) - 2v_{ik}] = 0. \tag{9.4}$$

That yields for the right side

$$\Delta \Phi_i - R_i{}^\alpha \Phi_\alpha = 0 \tag{9.5}$$

which is just our equation which has no solutions except zero.

Now we can integrate our Eqs. (9.3), reducing them to the self-adjoint differential expression $E(u_{ik})$, if we write (9.3) in the following form:

$$E(u_{ik}) = 2v_{ik} + \frac{\partial \psi_i}{\partial x_k} + \frac{\partial \psi_k}{\partial x_i} - \frac{\partial \psi^\alpha}{\partial x_\alpha} g_{ik} \tag{9.6}$$

where we have put

$$\psi_i = \sigma_i + \Phi_i \tag{9.7}$$

with

$$\sigma_i = \frac{\partial u_i^\alpha}{\partial x_\alpha} . \tag{9.8}$$

We can treat ψ_i as well as Φ_i as a free vector. Comparing the Eq. (9.6) with the form of the solution (3.18) obtained by using Mach's principle we observe the following connection between the vector ϕ_i appearing in (3.16), and the vector Ψ_i appearing in (9.6):

$$\psi_i = \Delta\phi_i + R_i^\alpha \phi_\alpha . \tag{9.9}$$

This can be proved easily if we introduce the solution (3.16) in the equation (9.6)[14]

The vanishing of Φ_i results in the following connection between the vector potential and the metrical quantities

$$\Delta\phi_i + R_i^\alpha \phi_\alpha = \frac{\partial u_i^\alpha}{\partial x_\alpha} . \tag{9.10}$$

In the case of our Hamiltonian we see that the right side *vanishes* for the expression (4.4) as well as for (4.6) if we employ our first integral $R = \text{const.}$

We obtain therefore the following fundamental equation as the determining equation of ϕ_i:

$$\Delta\phi_i + R_i^\alpha \phi_\alpha = 0. \tag{9.11}$$

This equation is the *exact equation for the vector potential.* We observe the familiar potential equation as a first approximation and we can also deduce immediately the conservation law, building

$$\frac{\partial}{\partial x_\alpha}(\Delta\phi^\alpha + R^{\alpha\sigma}\phi_\sigma) = 0 \tag{9.12}$$

that yields

$$\Delta \frac{\partial \phi^\alpha}{\partial x_\alpha} = 0 \tag{9.13}$$

and leads to the results (6.3) and (6.4).

We can consider the Eq. (9.11) also as a consequence of the conservation

[14] Unfortunately in the corresponding treatment of the previous paper (Zeits. f. Physik **32,** 169 (1925)) the author made a mistake, which was also recognized by Professor H. P. Robertson. In the equation of (31) and (32) of this paper there appears a minus sign instead of the correct formula:

$$\Phi_i = \tfrac{1}{2}(\Delta\Psi_i + R_i^\alpha \Psi_\alpha). \tag{32}$$

Having discovered this error already a long time ago, I could never understand the real significance of this change in the sign on comparing (32) in the rectified form with the Eq. (35) in which the minus sign is valid. In the present theory this change of the signs has a fundamental importance. While the equation with the minus sign is supposed to have no proper solutions besides zero, the equation with the plus sign is the fundamental equation for the vector potential and the existence of the material particles must be connected with the proper solutions of this equation. In fact just this change of signs causes Maxwell's equations to appear in *exact form* as shown in the next chapter.

law of R_{ik}, because the vanishing of the right side of (9.10), using for u_{ik} the form (4.4), is just a consequence of this law. We had originally 10+4 equations: the 10 field equations for R_{ik} and the vectorial conservation law. The latter conditions caused a Lagrangian factor ϕ_i which must be determined by the surplus equations. Indeed, we can consider the determining equation (9.11) as a representation for the conservation law. We have at last 10+4 equations for 10+4 quantities, the R_{ik} and the ϕ_i. In the first place there exist the 10 field equations in the form (7.1), which are to be completed only by the gravitational terms given by the second Hamiltonian H_2 and by the cosmological term. Secondly, there are the four Eqs. (9.11) for the vector potential.

We expect the possibility of proper solutions of (9.11) which must represent physically the material particles electron and proton. In the perception of this theory both electricity and gravitation are manifestation of a certain geometrical structure of the Universe. This structure is of the Riemannian type and governed by a Hamiltonian function which has the simplest form if we require the gauge-invariance.

Any completion of these equations by new "material" terms like the completion of Einstein's gravitational equations by a phenomenological matter-energy-tensor would be impossible. Such a completion means in fact the *denial* of the field equations in certain regions and is practically equivalent to the permission of singularities. The fundamental idea of this theory, however, is the consideration of Einstein's curvature-tensor R_{ik} as the fundamental characteristic quantity of Riemann's geometry. That is only possible if we exclude any singularities of the metric.

To determine the vector potential ϕ_i as a proper solution of an equation like (9.11) without further restrictions does not seem to be possible if we consider for example the arbitrariness of a factor in such a solution which does not correspond to the real physical statement. We must realize, however, that we should not conceive these equations as equations with definitely determined coefficients. There is a strict interconnection between vector potential and metric; on the one hand the metric is determined by the vector potential, on the other hand the constitutive equation of the vector potential is essentially influenced by the metric which produces the possibility of a proper solution. Thus the fundamental differential Eq. (9.11) is in reality a kind of complicated non-linear equation and the arbitrariness of a linear proper solution does not appear in the solution of such an equation. To determine the ultimate material particles and their inter-action as the proper solution of such a system of non-linear equations seems to me very satisfactory, especially in regard to such problems as the equality of the charge of all the electrons. The limitation attained by such a requirement without further restrictions seems to correspond to the real degree of limitation to be found in nature and corresponds also to a natural mathematical point of view according to which the solution of a variation problem is given by some differential equations completed by corresponding natural boundary conditions which can also become superfluous if the manifold is closed.

10. Connection with Maxwell's Equations and with the Completed form of Maxwell's Equations.

If we compare the left sides of (9.5) and (9.11) the change in the sign of the second term has a striking significance with respect to Maxwell's equations. The left side of (9.5) arises if we build the divergence of the *symmetric* combination

$$\frac{\partial \phi_i}{\partial x_k} + \frac{\partial \phi_k}{\partial x_i} \tag{10.1}$$

noticing the fact that we can leave out the additional term $-(\partial \phi^\alpha / \partial x_\alpha) g_{ik}$, which vanishes as a consequence of the conservation law of the vector potential. The left side of (9.11) arises if we build the divergence of the *antisymmetric* combination

$$F_{ik} = \frac{\partial \phi_i}{\partial x_k} - \frac{\partial \phi_k}{\partial x_i}. \tag{10.2}$$

The antisymmetric tensor F_{ik} must be considered as the electro-magnetic field strength. The equation

$$\operatorname{div} F_{ik} = 0 \tag{10.3}$$

is the one system of Maxwell's equations. The other system

$$\operatorname{div} F_{ik}^* = 0 \tag{10.4}$$

belonging to the "dual field-strength" F_{ik}^*, is an immediate consequence of (9.15). Thus *the whole system of Maxwell's equations is fulfilled without any change.* We may also build the symmetric Maxwellian stress-energy-tensor

$$S_{ik} = F_i{}^\alpha F_{k\alpha} - \tfrac{1}{4} F_{\alpha\rho} F^{\alpha\rho} g_{ik}. \tag{10.5}$$

The divergence of this tensor vanishes also exactly

$$\operatorname{div} S_{ik} = 0. \tag{10.6}$$

We obtain therefore, the surprising result that the classical form of electromagnetism appears in a very strict connection with Riemann's geometry. The difference compared with the former situation is that the electromagnetic quantities appear merely as *auxiliary quantities* in building up a metric. And the interaction between electromagnetic and metrical quantities is quite different from that assumed in any earlier attempt to combine electromagnetism with Einstein's theory of gravitation. As a consequence of this new interaction of metrical and electromagnetic quantities we may expect quite new insights in the problems connected with the constitution of matter.

However, a closer examination of the proper solutions of Maxwell's equations contradicts strongly what we expect. The equation

$$\frac{\partial (g)^{1/2} F^{91}}{\partial x_1} + \frac{\partial (g)^{1/2} F^{92}}{\partial x_2} + \frac{\partial (g)^{1/2} F^{93}}{\partial x_3} = 0 \tag{10.7}$$

in which the operation $\partial/\partial x_i$ now means an ordinary differentiation, permits the application of Gauss' theorem and leads generally to the *vanishing* of the electric charge.[15] We see, therefore, that the introduction of a curved manifold is not sufficient for producing useful proper solutions of the Maxwellian equations.

If we review our method of building up the fundamental equations, we observe the use of some restrictional assumptions which are in reality not necessary for the fulfilling of the field equations. We introduced the vector potential ϕ, by the use of a process which we called "Mach's principle." During this process we made the assumption that the equation $E(u_{ik})=0$ as well as the equation $\Delta\Phi=0$ should not have nontrivial proper solutions. The latter restriction led to the first integrals $R=$ const. and div $\phi_1=0$.

In Chapter 9 we proved the validity of the integrated form of the field equations without the help of Mach's principle. It turned out that we have to add only the equation (9.10) as a condition for the vector potential. Any other condition is unnecessary. Since we have seen now that the strict validity of Maxwell's equations leads to a contradiction, we have to abandon the unnecessary supposition that (5.4) shall be the only solution of (5.3). And we have therefore to abandon our first integrals (5.4) and (6.4). These integrals are responsible for the strict appearance of Maxwell's equations. If we do not make the above-mentioned restriction, we get by (9.10) the following constitutive equation for the vector potential ϕ_1:

$$\Delta\phi_i + R_i{}^{\alpha}\phi_{\alpha} = (1 + 2C)\frac{\partial R}{\partial x_i} \tag{10.8}$$

and Maxwell's equations for the antisymmetric field strngth F_{ik}, defined again by (10.2), appear in the following form:

$$\frac{\partial F_i{}^{\alpha}}{\partial x_{\alpha}} + \frac{\partial \phi}{\partial x_i} = 0 \tag{10.9}$$

where we introduced the scalar ϕ by putting

$$\phi = \frac{\partial \phi^{\alpha}}{\partial x_{\alpha}} - (1 + 2c)R. \tag{10.10}$$

The second set of Maxwell's equations remains unchanged in the previous form (10.4).

If we use the familiar notations of the 3-dimensional vector-analysis, we can write our equations in the case of a Euclidian metric in the following form

[15] The equation (10.7) being valid without exception also in the central part of the particle, we can build a surface integral over a closed surface surrounding the particle at such a distance that the metric is there practically Euclidian. That gives

$$\int F_{\Phi\alpha} v_{\alpha} dS = 0$$

The left integral, built in the rest system of the particle, has the significance of the charge.

$$\frac{1}{ic}\frac{\partial F}{\partial t} - \text{curl } F - \text{grad } \Phi = 0$$
$$\text{div } F + \frac{\partial \Phi}{ic\partial t} = 0. \tag{10.11}$$

We introduced there the complex vector F with the significance

$$F = H + iE \tag{10.12}$$

where H is the magnetic, E the electric field-strength. Compared with Maxwell's equations this system is enlared with the terms containing Φ. This self-adjoint system of equations has just the right degree of determination, constituting 4 equations for the 4 quantities F and ϕ. Generally we have to consider F and Φ as complex quantities. The system is then equivalent—as I have proved in a former investigation[16]— to Dirac's equation for the electron, if we leave out the term with the mass. The only difference in our case is that Φ must be real.

The stress-energy-tensor of Maxwell must be enlarged in the following form:[17]

$$S_{ik} = F_i{}^{\alpha}F_{k\alpha} - \tfrac{1}{4}(F_{\alpha\beta}F^{\alpha\beta} + 2\Phi^2)g_{ik} + \Phi F_{ik} \tag{10.13}$$

in order that the divergence vanish. This enlarged tensor is no longer symmetric, due to the last term which is antisymmetric.

The difficulty of the vanishing of the charge is now overcome. We see, however, that the equation (10.7) appears also now if we require a *static* solution. Because the equation $\Delta\Phi = 0$ has certainly no *statical* proper solutions. It seems that there is a general *pulsation* in the metrical quantities which gives a "guiding field" for the pulsation of all matter. It seems that we must understand the electrostatic field as a *quadratic effect* of these pulsating fields, taking into consideration the quadratic character of our action-principle. This character appears also in the expression of the ponderomotive force as a product of charge and field-strength.

[16] Zeits. f. Physik **57**, 447 (1929).
[17] Zeits. f. Physik **57**, (1929), Eq. (15′).

VELOCITY OF SOUND IN TUBES AT AUDIBLE AND ULTRASONIC FREQUENCIES

By Charles B. Vance
Indiana University

(Received November 9, 1931)

Abstract

The velocity of sound in air contained in glass tubes ranging in diameter from 0.1 cm to 3.0 cm was measured at ultrasonic frequencies over a frequency range from 30 to 200 kilocycles per second by using quartz crystals as sound oscillators, and the Kundt's dust tube method. The results do not agree with those calculated by use of Helmholtz-Kirchhoff equation

$$V^1 = Vo(1 - C/2R(2\pi N)^{1/2}).$$

An equation of the form

$$V^1 = Vo(1 - C/D^2 - K/D(N)^{1/2})$$

is fairly satisfactory for both audible and ultrasonic frequencies. In this equation $V_0 = 331.77$ m/sec., $C = 0.001512$, $K = 0.174$, N = frequency and D = diameter of the tube. The agreement between experimental and calculated values is better for the larger tubes than for the smaller.

Introduction

EXPERIMENTAL observations of the velocity of sound in pipes may be divided into three classes, in which direct measurements were made of the time required (1) for the sound to pass from one end of the pipe to the other end; (2) to return to the source after being reflected from the closed farther end; (3) to traverse a U-tube and return to the source.

From experiments Regnault[1] found that the velocity of sound increased with the diameter of the pipes, reaching the ordinary value in large pipes. Rink's[2] analysis of Regnault's observations gives a mean value of the velocity of 330.5 m/sec. at 0°C. Voille and Vautier,[3] by observations similar to those of Regnault, obtained 331.1 m/sec. at 0°C.

Helmholtz[4] and Kirchhoff[5] considered theoretically the propagation of sound waves in small tubes and arrived at the following expression for the speed:

$$V^1 = V_0(1 - C/2R(2\pi N)^{1/2}) \qquad (1)$$

where V^1 is the velocity of sound for frequency N, in a pipe of radius R, V_0 being the velocity in free air at 0°C and C a constant with no definite agreement as to its exact value. Schneebeli[6] and Seebeck[7] found from their experiments that the decrease of velocity varies as the radius of the tube but that

[1] V. Regnault, Comptes Rendus **66,** 209 (1868).
[2] Rink, Pogg. Ann. **149,** 533 (1873).
[3] Voille and Vautier, Phil. Mag. **26,** 77 (1888).
[4] Helmholtz, Crelles Jour. **57,** 1 (1859).
[5] Kirchhoff, Pogg. Ann. **134,** 177 (1868).
[6] Schneebeli, Pogg. Ann. **136,** 296 (1869).
[7] Seebeck, Pogg. Ann. **139,** 104 (1870).

when the frequency varies it is inversely proportional to the squareroot of the cube of the frequency instead of the squareroot as given in Eq. (1). Müller[8] found that the equation is not valid but that the velocity of sound in pipes depends on the material of the pipes. Stevens[9] found that C is valid for pipes of 2.0 cm to 4.0 cm diameters, while Schulze[10] claimed that C depends on the diameter and nature of the pipes. The results obtained by Wertheim[11] and Blaikley[12] are in support of Eq. (1).

The above results were obtained with frequencies (256 to 5550) within the audible region. When the frequency of the sound is above the audible limit, or too high to be estimated by the ear, the velocity is more conveniently measured by Rayleigh's[13] "stationary wave" method or Kundt's[14] dust tube. W. Altberg[15] measured the wave-length of the sound emitted by an electric spark, frequency $N=340{,}000$, and found that the product $N\lambda$ was in moderately good agreement with the accepted values for the velocity of sound in air. A. Campbell and D. W. Dye[16] obtained a similar result at $N=900{,}000$ using a high frequency spark as a source and a Kundt's dust tube to indicate the wave-length. A. L. Foley[17] found by the photographic method that a very intense sound, such as produced by an intense electric spark, has an abnormally high velocity at points near the source where the intensity is great. If such a source be placed in one end of a pipe so that the wave does not have any chance to expand and thus decrease in intensity, such a wave travels through a tube with a higher velocity than normal. If the spark were placed at a distance of two to five centimeters, depending on the intensity of the spark, from the end of the tube, the velocity in the tube would be less than it is in free air.

In recent years the piezoelectic property of crystals has been used as a source of high frequency vibrations. Cady[18] made a thorough investigation of crystal oscillations and crystal resonators, and adapted them to use as constants of electrical frequency. Harrison[19] and others made an investigation as to the possible modes of vibration of quartz crystals. Pierce,[20] Reid,[21] Wood and Loomis,[22] Hubbard and Loomis[23] and others have used piezoelec-

[8] Müller, Ann. d. Physik **11,** 331 (1903).
[9] Stevens, Ann. d. Physik **7,** 285, (1902).
[10] Schulze, Ann. d. Physik **13,** 1060 (1904).
[11] Wertheim, Pogg. Ann. **77,** 427 (1844).
[12] Blaikley, Phil. Mag. **16,** 477 (1883).
[13] Rayleigh, Sound **2,** 403.
[14] Kundt, Pogg. Ann. **135,** 337 (1868).
[15] W. Altberg, Ann. d. Physik **23,** (1907).
[16] A. Campbell and D. W. Dye, Electrician **66,** 862 (1911).
[17] A. L. Foley, Phys. Rev. **14,** 2 (1919).
[18] W. G. Cady, Proc. Inst. Rad. Eng. **10,** 83 (1922); Phys. Rev. **19,** 1 (1922); Jour. Opt. Soc. Am. **10,** No. 4, April (1925).
[19] J. R. Harrison, Proc. Inst. Rad. Eng. 1040, Dec. 1927.
[20] G. W. Pierce, Proc. Am. Acad. of Arts and Science **60,** 271 (1925).
[21] C. D. Reid, Phys. Rev. **35,** 814 (1930).
[22] R. W. Wood and A. L. Loomis, Phil. Mag. **4,** 417 (1927).
[23] J. C. Hubbard and A. L. Loomis, Phil. Mag. **5,** 1177 (1928).

tric crystals as the source of ultrasonic vibrations, of accurately known frequency, for the determination of the velocity of sound in liquids or gases.

Experimental Procedure

The velocity of sound waves in tubes of small diameters has been found in most cases by the Kundt's dust tube or some resonance tube method. It was suggested that quartz resonators be used as a source of high frequency sound for the investigation of sound waves in tubes of small diameter to find whether or not the Helmholtz-Kirchhoff equation is valid for ultrasonic waves. This work was carried on by the writer in the ultrasonic region at frequencies from 30 kc to 200 kc, using glass tubes varying in diameter from 0.1 cm to 3.0 cm. A quartz crystal was mounted as shown in Fig. 1 which is a modified form of Pierce's circuit.

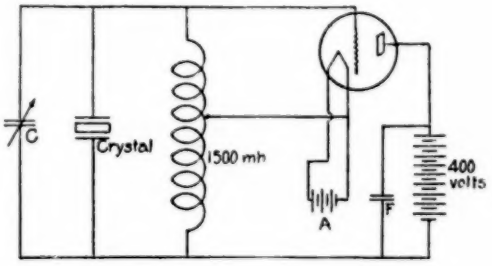


Fig. 1. Diagram of apparatus.

A type UX210 Radiotron tube was used with 400 volts on the plate. The capacity and inductance used depended on the frequency of the various crystals (7 crystals were used). When the oscillating circuit was tuned to crystal frequency, the crystal was set in vigorous vibration, producing ultrasonic waves of constant frequency. Round glass tubes were fitted with movable pistons and lycopodium powder was dusted in them. Each tube in turn was brought close to the face of the vibrating crystal and the position of the piston adjusted so as to produce nodes of lycopodium dust. The distance between nodes was then measured and this value taken as the half wave-length. Knowing this value and the frequency of the crystal, the velocity of the sound wave in the tube was calculated from $V = N\lambda$, where V is the speed, N the frequency, and λ the wave-length. The crystals which were calibrated by the Bureau of Standards had the following frequencies: 30.3 kc, 49.58 kc, 51.5 kc, 69.4 kc, 100 kc, 145 kc, and 200 kc. In order to obtain the best possible results, care was taken to avoid change in the frequency of the crystals. This was done (1) by having the crystal perfectly clean, (2) by proper adjustment of the electrodes to the crystal, thus preventing sparking and heating, (3) by operating the crystal for a comparatively short time only, to avoid heating. The lycopodium powder and tubes were kept dry, as any dampness in either of these made results difficult to obtain.

Discussion

For measuring the distance between nodes a micrometer microscope, which could be read accurately to 0.005 cm, was used. The maximum point on the nodes could not be located accurately but after a number of trials, the probable error of any one setting of the micrometer was found to be less than 0.5 mm. Then in order to increase the accuracy of the readings the microscope was set on every tenth node and the reading taken. The difference between any two consecutive readings was divided by ten, thus reducing the error to one-tenth of that for one reading alone. The value of the wavelength for each tube and crystal was obtained by taking the average of fifty or more measurements, as described above, and then multiplying this value by two. Due to the difficulty in obtaining distinct nodes beyond the thirty-fifth in the smaller tubes, the measurements were limited to the first thirty-five nodes in all tubes. The first two or three nodes were not measured as they were more or less distorted. The temperature at which the observations were made was 21°C, but the velocity of sound in every case was reduced to 0°C for comparison. In reducing the velocity to 0°C it was assumed that the velocity of ultrasonic waves varies with the temperature the same as audible waves, i.e., 0.6 m/sec. per degree centigrade change in temperature.

No correction was made for the humidity. The relative humidity was less than 30 percent and was practically constant. Wood[24] states that the calculated velocity in air saturated with moisture at 10° C is from 2 to 3 feet/sec. greater than in dry air. Stewart-Lindsay[25] states that the velocity of sound in saturated air at 20° C and 760 mm Hg is about 0.33 percent greater than that in perfectly dry air at the same temperature and pressure.

The velocity of sound in tubes of different diameters was calculated by use of the Helmholtz-Kirchhoff equation in which V_0 is given the value of 331.7 7m/sec., and C that of 0.132, which is the ratio of the coefficient of viscosity of air to its density at 0°C, or what Richardson[26] calls the kinematic coefficient of viscosity.

Table I shows my experimental and calculated results.

The experimental results are shown graphically in Fig. 2, in which the average value of the velocities for all frequencies for a given tube diameter is plotted against the tube diameter.

It will be seen that the values calculated by use of Eq. (1) do not agree with the experimental values (see Table I). A greater decrease would be expected for the smaller tubes due to the increased effect of viscosity and heat conduction. When the diameter of the tube is small the conduction of heat from the center of the air column to the walls becomes more and more rapid.[27] The velocity of sound in such a tube might be expected to be more nearly that of Newton's isothermal value $K/\rho^{1/2}$. Viscosity may be expected to exert a greater influence when sound waves pass through a relatively narrow tube

[24] A. B. Wood, A Textbook of Sound, p .231.
[25] Stewart-Lindsay, Acoustics, p. 308.
[26] F. G. Richardson, Sound, p. 166.

TABLE I. *Velocity of sound in glass tubes, of different diameters, and at different frequencies.*

Frequency	Diameter of tube	Velocity experimental at 0°C	Velocity from Eq. (1) at 0°C	Velocity from Eq. (3) at 0°C	Difference exp−Eq. (3)
30.3 kc	3.0 cm	332.30 m/sec.	331.74 m/sec.	331.47 m/sec.	+1.83
	1.0	332.05	331.67	329.76	+2.89
	0.7	328.55	331.60	327.85	+0.70
	0.5	325.05	331.56	324.37	+0.68
	0.4	319.90	331.51	320.46	−0.56
	0.3	310.68	331.42	311.86	−1.18
	0.2	288.49	331.28	287.64	+0.85
	0.1	161.75	330.75	158.59	+3.16
49.58 kc	3.0	331.88	331.74	331.50	+0.38
	1.0	331.70	331.68	329.78	+1.92
	0.7	328.29	331.62	327.92	+0.37
	0.5	324.70	331.59	324.47	+0.23
	0.4	319.55	331.55	320.66	−1.11
	0.3	310.35	331.48	312.20	−1.85
	0.2	288.28	331.33	288.14	+0.14
	0.1	161.53	330.91	159.25	+2.28
51.5 kc	3.0	331.82	331.74	331.50	+0.32
	1.0	331.68	331.69	329.81	+1.87
	0.7	328.26	331.65	327.99	+0.27
	0.5	324.66	331.62	324.47	+0.19
	0.4	319.50	331.58	320.57	−1.09
	0.3	310.29	331.52	312.20	−1.91
	0.2	288.24	331.39	288.14	+0.10
	0.1	161.51	331.01	159.58	+1.93
69.4 kc	3.0	331.66	331.75	331.50	+0.16
	1.0	331.46	331.70	329.85	+1.61
	0.7	328.07	331.66	328.05	+0.02
	0.5	324.48	331.64	324.54	−0.06
	0.4	319.30	331.61	320.69	−1.39
	0.3	310.08	331.55	312.26	−2.18
	0.2	288.11	331.44	288.31	−0.21
	0.1	161.40	331.11	159.71	+1.69
100 kc	3.0	331.40	331.75	331.50	−0.10
	1.0	331.20	331.71	329.88	+1.32
	0.7	327.90	331.69	328.12	−0.32
	0.5	324.32	331.67	324.64	−0.32
	0.4	319.17	331.63	320.79	−1.62
	0.3	309.90	331.57	312.36	−2.46
	0.2	288.02	331.50	288.47	−0.45
	0.1	161.30	331.21	160.08	+1.22
145 kc	3.0	331.35	331.75	331.54	−0.19
	1.0	331.10	331.72	329.91	+1.19
	0.7	327.80	331.70	328.12	−0.42
	0.5	324.18	331.68	324.70	−0.52
	0.4	319.02	331.65	320.85	−1.83
	0.3	309.78	331.62	312.49	−2.71
	0.2	288.01	331.54	288.64	−0.63
	0.1	161.28	331.31	160.38	+0.90
200 kc	3.0	331.20	331.76	331.54	−0.34
	1.0	331.00	331.73	329.95	+1.05
	0.7	327.65	331.71	328.15	−0.50
	0.5	324.13	331.69	324.74	−0.61
	0.4	318.97	331.67	320.92	−1.95
	0.3	309.73	331.64	312.56	−2.83
	0.2	287.98	331.57	288.97	−0.99
	0.1	161.25	331.38	160.58	+0.67

in which the wall offers greater resistance to the motion of the air in contact with it.[28]

When the relation between the wave-length at the frequencies used, 30 kc to 200 kc, and the diameter of the tubes (smallest 0.1 cm diameter) is taken into consideration as compared with the frequency and diameter of tubes at audible frequency, it is found that the diameters of the tubes used are relatively large. The above experimental results, Table I, show that in the ultrasonic region (30 kc to 200 kc) the frequency has but little effect on the velocity of sound. The percent of difference between the velocity of sound at any frequency and the average velocity of sound for all frequencies for a given tube diameter is less than 0.2 of one percent.

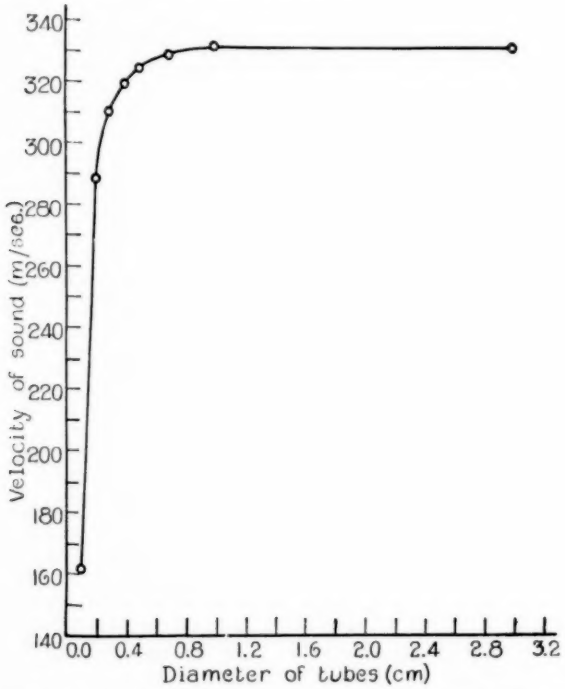


Fig. 2. Relation between the velocity of sound in tubes and the diameter of tubes.

The effect of viscosity and heat conduction in reducing the sound energy is greatly increased when a gaseous medium is brought into contact with a large surface area of a solid or a liquid. Lord Rayleigh[29] shows that the extent to which the 'viscous' dragging effect penetrates into the gas is proportional to $\nu/D(N)^{1/2}$. The effect of viscosity in modifying the motion is dependent on the ratio of $\mu/\rho=\nu$ rather than on μ, the viscosity, alone. Kirchhoff pointed out that the loss of energy due to heat conduction in a medium cannot be neglected. If the diameter if the tube is large compared with the wave-

[27] A. B. Wood, A Textbook of Sound, p. 323.
[28] Stewart-Lindsay, Acoustics, p. 68.
[29] Lord Rayleigh, Sound **2**, p. 319.

length the effect of viscosity will be small while if the tube diameter is small compared with the wave-length the effect will be much greater.

Foley[17] has shown that the velocity of sound in a pipe depends on the intensity of the wave as it enters the pipe rather than upon the intensity of the sound source. The source of sound being outside the tubes the amount of energy entering the tubes, placed the same distance from the sound source, would be directly proportional to the square of the diameter of the tube.

TABLE II.

Observer	Frequency	Diameter of tube	Velocity experimental at 0°C	Velocity from Eq. (3) at 0°C	Difference exp−Eq. (3)
Low[30]	256	0.93 cm	320.60 m/sec.	325.80 m/sec.	−5.20
		1.71	325.24	327.19	−1.95
		2.80	327.29	327.56	+0.27
Seebeck[7]	320	0.34	317.26	309.38	+8.90
		0.90	328.02	326.03	+1.99
		1.75	329.24	327.46	+1.78
Schulze[10]	384	0.101	258.00	133.44	+124.56
		0.151	282.00	235.56	+46.44
Schulze[10]	512	0.101	265.00	139.34	+125.66
Seebeck[7]	512	0.34	322.98	310.39	+12.79
Seebeck[7]	512	0.90	328.44	326.76	+1.68
Seebeck[7]	512	1.75	330.92	328.29	+2.63
Low[30]	512	0.93	323.60	326.86	−3.26
Low[30]	512	1.71	326.70	328.25	−1.55
Low[30]	512	2.80	328.33	328.62	−0.29
Muller[8]	903	1.55	327.30	328.88	−1.58
Low[30]	1023	0.93	325.29	327.69	−2.40
		1.71	327.80	329.04	−1.24
		2.80	328.68	329.45	−0.77
Muller[8]	2482	1.55	330.20	329.75	+0.45
Kundt[14]	3700	0.35	318.88	317.07	+1.81
		0.65	327.14	326.69	+0.45
		1.30	329.88	329.65	+0.23
Kundt[14]	5550	0.65	328.14	326.89	+1.25
		1.30	330.87	329.85	+1.02

In order to include the effect of viscosity, heat conduction and intensity for all relations of wave-length to tube diameters, an equation of the form

$$V^1 = V_0(1 - C/D^n - K/D(N)^{1/2}) \qquad (2)$$

was used with V_0 as the velocity of sound at 0°C, V^1 as the velocity of sound in a tube of diameter D and at a frequency N, and C, K and n were constants. The value of C, K and n were determined and found to be $C=0.00512$, $K=0.174$, and $n=2$. It was found that the best value for the velocity of sound at audible frequencies and 0°C is 331.45 m/sec.,[31] while for ultrasonic frequen-

[30] Low, Ann. d. Physik **52**, 841 (1894).

[31] A. L. Foley, International Critical Tables **6**.

cies and 0°C it appears to be 332.08 m/sec.[32] The value for V_0 was taken as the average of these two values or 331.77 m/sec. Eq. (2) now becomes

$$V^1 = V_0(1 - 0.00512/D^2 - 0.174/D(N)^{1/2}). \quad (3)$$

At the various ultrasonic frequencies the velocity of sound in tubes of different diameters was calculated by use of Eq. (3) and the results thus obtained were compared with my experimental values as shown in Table I. At various audible frequencies the velocity of sound in tubes of different diameters was calculated by use of Eq. (3). The results were compared with the experimental values obtained by other experimenters using glass tubes. Table II shows the comparison. From Table I we see that for ultrasonic frequencies, the experimental and calculated values for the velocity of sound agree very well, while from Table II for audible frequencies, the experimental and calculated values for the velocity of sound compare fairly well except for the smallest tubes, 0.101 and 0.151 cm diameters. As the experimental and calculated values for the velocity of sound for these tubes differ widely, I concluded to redetermine them at the same frequencies and in tubes of approximately the same diameter as were used by Schulze.

My method was as follows: the sound produced by a constant frequency phonograph record was amplified and then transmitted to a telephone receiver held close to the end of the tube. In making these measurements the same general plan was followed as that used for the measurements at ultrasonic frequency. The results (the average of 20 measurements for each tube and frequency) thus obtained are shown in Table III.

TABLE III.

Frequency	Diameter of tube	Velocity experimental at 0°C	Velocity from Eq. (3) at 0°C	Difference exp−Eq. (3)
384	0.105 cm	151.04 m/sec.	149.30 m/sec.	+1.74
	0.160	250.72	247.17	+3.55
512	0.105	153.60	152.28	+1.32
	0.160	251.80	249.48	+2.31

It will be seen by Table III that my values are less than those obtained by Schulze but that they agree fairly well with those calculated by use of Eq. (3).

I wish to express my appreciation to Dr. A. L. Foley for his very valuable suggestions, and also to Dr. R. R. Ramsey for his interest in the work.

[32] C. D. Reid, Phys. Rev. **37**, 1147 (1931).

LETTERS TO THE EDITOR

Prompt publication of brief reports of important discoveries in physics may be secured by addressing them to this department. Closing dates for this department are, for the first issue of the month, the twenty-eighth of the preceding month; for the second issue, the thirteenth of the month. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents.

Increasing the Charge Sensitivity of Vacuum Tube Amplifiers

For several years various systems of neutralizing the input capacity of vacuum tube amplifiers[1] have been known. These systems have not been widely used nor well understood. For this reason I wish to show the possibilities for their use in increasing the charge sensitivity of vacuum tube amplifiers and also the speed of their operation.

The capacity of the input circuit of a vacuum tube limits the charge sensitivity to the order of 10^{-15} to 10^{-16} coulomb per mm, so that, with a current of 10^{-17} amperes, it takes from 10 seconds to 2 minutes to produce a readable deflection. One of the fundamental circuits for input capacity neutralization is shown in Fig. 1.

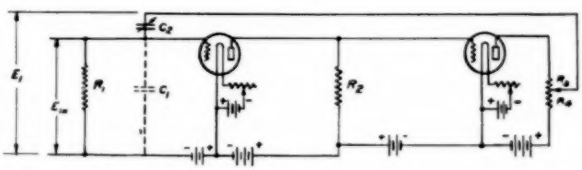


Fig. 1. Input capacity neutralization for low current measurement.

The conditions for complete neutralization are that C_1 equals C_2 and the ratio of R_3/R_4 is such that E_1 equals $2E_{1N}$. These conditions, of course, are not stable and represent the point at which oscillations start. The amount of capacity neutralization which is possible depends on the stability of the circuit. Preliminary measurements indicate that good operation is possible with the effective input capacity reduced to 1/20 of its original value. It is felt that by careful selection of circuits and tubes this value might be increased at least tenfold.

Charge sensitivites of 2×10^{-17} coulomb/mm have been observed with the FP-54 Pliotron tube.[2,3] By using the above method, this sensitivity might be increased to 2×10^{-19} coulomb/mm, or approximately one electron per scale division, and still retain good stability.

G. F. METCALF
Vacuum Tube Engineering Department,
General Electric Company,
Schenectady, New York,
January 21, 1932.

[1] Brillouin, Patent **1,** 404, 574, January 24, 1922.

[2] Metcalf and Thompson, Phys. Rev. **36,** 1489–1494 (1930).

[3] L. A. DuBridge, Phys. Rev. **37,** 392–400 (1931).

The Effect of Heat on Mercury Bands

It is well known from the work of Niewodniczanski (Zeits. f. Physik **49,** 59 (1928)) and others that moderate heating destroys the 4850 band and enhances the 3300 band in fluorescence and strong heating destroys also the 3300 band. This effect and the effect of heat on other mercury bands may be conveniently observed in a Tesla coil electrode-

less discharge. A greatly reduced image of a 15 cm discharge tube was focussed for λ2540 on the slit of a Hilger E31 quartz spectrograph, and the spectrum photographed while

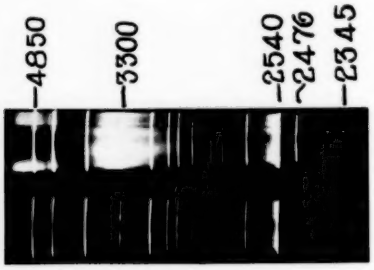


Fig. 1.

the center of the discharge tube was heated by a blow torch. The spectrum for two exposure times is shown in Fig. 1. The difference between the center and edges represents the effect produced by heat. The pressure may be considered to be the same for all portions of the tube.

Heating is observed to destroy the 4850 band and to greatly reduce the intensities of the bands at 2345, 2476, 2540, and the continuous spectrum between 2536 and 3300. Moderate heating enhances 3300 but strong heating destroys it. The majority of the mercury arc lines are unaffected by heat but 5460, 4358, 4047, and 4077 are noticeably weakened.

The fact that most of the mercury arc lines are not affected by heat indicates that the discharge conditions are nearly the same in the heated and unheated portions of the tube. The weakening of the bands must be due to the destruction of mercury molecules by heat. The weakening of the lines probably results from collisions of the first kind which carry excited atoms to higher energy states.

J. Gibson Winans
University of Wisconsin,
January 25, 1932.

The Resultant Electric Moment of Complex Molecules

The purpose here is to express the average resultant dipole moment in a form which may be readily evaluated provided that the potential energy associated with distortion of the molecule is known. The individual moments are assumed constant. A molecule may be idealized as a framework of lines along the valence bonds connecting the atoms. About certain of these bonds there is rotation. The experimental electric moment may be regarded as the vector sum of a series of electric moments lying along each of these lines. The moment along some lines may be zero. Let us begin with some end line and number them consecutively 1, 2, 3, etc., along the longest chain until another end line is reached. At points where there is branching the numbering along the different lines is to be distinguished by primes. Now consider a set of unit vectors lying along these lines with their directions pointing toward increasing numbers. Let θ_j be the angle which two consecutive vectors $j-1$ and j make with each other, being zero when they point in the same direction. Let us choose sets of right-handed rectangular coordinate systems in the following way. Each unit vector j coincides in position and direction with the x_j axis of a coordinate system. y_j lies in the plane determined by x_j and x_{j-1} in such a way that the angle between the positive direction of y_j and the negative direction of x_{j-1} is acute. A right-hand screw progressing along positive x_{j-1} turns through an angle, θ_j, in passing from positive z_{j-1} to positive z_j. The transformation A_j, which when applied to the components of a vector referred to the j coordinates transforms it to the $j-1$ coordinate system, is:

$$A_j = \begin{vmatrix} \cos\theta_j & -\sin\theta_j & 0 \\ \sin\theta_j\cos\theta_j & \cos\theta_j\cos\phi_j & -\sin\phi_j \\ \sin\theta_j\sin\phi_j & \cos\theta_j\sin\phi_j & \cos\phi_j \end{vmatrix}. \quad (1)$$

By repeated transformation each electric moment $c_{xj}e_{xj}$ referred to the j coordinates can be referred to a particular set of coordinates, say the set of index 1. e_{xj} is the unit vector in the x_j direction and c_{xj} gives the magnitude and sign of the electric moment in the x_j direction. The above method of specifying coordinates does not fix ϕ_2, which is taken to be equal to zero, which then determines A_2. The resultant vector sum of the n electric moments referred to coordinate system 1 is then

$$C_{x1}e_{x1} + C_{y1}e_{y1} + C_{z1}e_{z1} = \sum_{j=1}^{n} \prod_{k=2}^{j} A_k c_{xj} e_{xj} \quad (2)$$

k and j include primed suffixes where there are side chains. The term $j=1$ in the summation (2) is to mean $c_{x1}e_{x1}$.

If the length of the resultant vector is C, then, taking the scalar product, one obtains:

$$\begin{aligned} C^2 &= C^2_{x1} + C^2_{y1} + C^2_{z1} \\ &= \left(\left(\sum_{j=1}^{n} \prod_{k=2}^{j} A_k c_{xj} e_{xj}\right), \left(\sum_{j=1}^{n} \prod_{k=2}^{j} A_k c_{xj} e_{xj}\right)\right) \\ &= \left(\left(\sum_{j=1}^{n} c_{xj} e_{xj}\right), \right. \\ &\quad \left. \left(\sum_{j=1}^{n} \prod_{k=s}^{2} A'_k \sum_{j=1}^{n} \prod_{k=2}^{j} A_k c_{xj} e_{xj}\right)\right) \\ &= \sum_{j=1}^{n} c^2_{xs} \\ &\quad + 2 \sum_{j=1}^{n} \sum_{s<j} [(c_{xj} e_{xj}), (A'_j \cdots A'_s c_{xs} e_{xs})] \end{aligned} \tag{3}$$

A_j is the reverse transformation to A_j and is obtained from A_j by transposing rows into columns. To avoid a more complicated notation the result given in (3) is for a straight chain. The general result independent of the number of branching chains, may be stated. The square of the resultant electric moment is the sum of the squares of the individual moments plus twice all the possible scalar products of all the individual moments after one of these moments has been transformed by consecutive transformations along the reference frame to the coordinate system of the other. Having obtained a resultant for a set of vectors in a given configuration we now proceed to evaluate the average of this resultant weighted with the frequency with which the various configurations occur. The average of C^2 will be called C^2_a. The averaging is accomplished by evaluating the quantity:

$$C^2_a = \frac{\int c^2 e^{-E/kT} \sin\theta_2 d\theta_2 \cdots \sin\theta_m d\theta_m d\phi_s \cdots d\phi_m}{\int e^{-E/kT} \sin\theta_2 d\theta_2 \cdots \sin\theta_m d\theta_m d\phi_2 \cdots d\phi_m} \tag{4}$$

In certain cases (4) must be replaced by a summation over discrete energy levels. E is the energy expressed as a function of all the angles. The angles have indices running between 2 and m. Each θ is integrated between the limits 0 and π and each ϕ between the limits 0 and 2π.

Certain special cases are of interest. If E can be broken up into a sum of potential functions each of which only involves a θ_k and ϕ_k with the same index k, then the averaging can be carried out on the elements of each matrix before multiplying the matrices. A particularly simple case is that for which there is a very sharp minimum in E for a particular value of each angle θ and where E is independent of every angle ϕ. Then A'_k becomes, after averaging over all the angles

$$A'_k = \begin{vmatrix} \cos\theta_k & 0 & 0 \\ \sin\theta_k & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \tag{5}$$

where θ_k is now the angle corresponding to the minimum in the potential energy. Eq. (3) then becomes

$$C^2_a = \sum_{j=1}^{n} c^2_{xj} + 2 \sum_{j=1}^{n} \sum_{s<j} \prod_{k=j}^{s+1} \cos\theta_k c_{xj} c_{xs} \tag{6}$$

Eq. (6) may be stated in words. For free rotation about connecting lines the mean square sum of n vectors is equal to the sum of the squares of the lengths of the separate vectors, plus twice all the possible products of the lengths of two vectors multiplied by the product of the cosines of the angles θ_k made by the directed lines connecting the pair. θ_k is zero when the directed lines are continuations of each other. The product, $c_j c_s$, of the lengths of the vectors is positive if they point in the same direction in passing along a chain.

In organic molecules where stiff structures such as rings or double bonds occur, a frame may be chosen consisting simply of the straight lines about which there is free rotation. A group of moments inside the stiff structure is represented by a resultant vector passing through the nearest point of rotation and is essentially a side chain of one link. If free rotation is assumed (6) is then immediately applicable to this simplified framework. Every molecule is a special case of either (3) or (6).

These equations apply equally well if the vectors are the distances between atoms instead of the electric moments. For example assuming free rotation the root mean square distance, C_a, between the first and $(n+1)$st carbon atoms of a paraffin chain is

$$C_a = c_1(n + 2(n-1)\cos\theta + 2(n-2)\cos^2\theta \cdots + 2\cos^{n-1}\theta)^{1/2}$$

where c_1 is the distance between neighboring carbon atoms and where the angles between the successive bonds are zero for a straight chain.

These relationships are being used in connection with approximate calculations of potential energies in the interpretation of the observed lengths and electric moments of organic molecules.

HENRY EYRING
Frick Chemical Laboratory,
Princeton University,
Princeton, N. J.,
January 29, 1932.

Another New X-ray Non-Diagram Line

In a recent letter to the Editor (Phys. Rev. Feb. 1, 1932) the writer gave values for the wave-lengths of a non-diagram line found in the K spectrum of molybdenum, rhodium, palladium, and silver. The line was provisionally christened β_4 and its possible origin indicated.

The search for faint lines has been continued. Close to β_1 on the short wave-length side another and somewhat fainter line has been found and will be referred to as β_5. The wave-length was determined on a double-crystal spectrometer working in the second order. The International Critical Table values of λ for $\gamma(\beta_2)$ and β_1 were accepted and the wave-length of β_5 found by interpolation. For convenience in determining the probable origin of the new line, its frequency was computed as on page 35, Vol. VI, International Critical Tables. The frequency of the new line was then compared with differences of limiting frequencies of x-ray levels as given in the tables.

The line may be explained as the result of either of two equivalent double transitions, i.e., $(M_{21}\rightarrow K)+(M_{22}\rightarrow M_{11})$ or $(M_{22}\rightarrow K)+(M_{21}\rightarrow M_{11})$. Both demand a double ionization resulting from the simultaneous removal of a K and an M_{11} electron and both demand an improbable transition from one M level to another, but a line less than one one thousandth as intense as β_1 must have something very improbable in its origin.

	$\lambda\beta_5$(X.U.)	$\nu\beta_5$	$(\nu K \lim - \nu M_{21}) + (\nu M_{11} - \nu M_{22})$
Mo	629.46	1448.4	1448.5
Rh	542.87	1679.4	1679.6
Pd	517.84	1760.6	1760.1
Ag	494.27	1844.7	1844.3

Two groups of even fainter lines have been detected. One group lies at the foot of γ on the long wave-length side. The other group lies at the foot of β_2 on the long wave-length side. Each group seems to consist of about five overlapping lines. The present rather uncertain determinations of their wave-lengths indicate that they cannot be explained by any combination of transitions between known x-ray energy levels.

P. A. Ross
Department of Physics,
Stanford University,
California,
January 30, 1932.

Note on the New Effect Produced by Action of X-rays on Matter

G. I. Pokrowski[1] reported in this journal the results of a series of experiments to determine what effect x-ray irradiation of metals may have. These experiments seem to show that when elements of high atomic number are irradiated with x-rays, they acquire feeble radioactive properties. Two methods were used to demonstrate this feeble radioactivity; first, the ionization produced by the irradiated samples, and second, the scintillations produced on a screen of zinc blende placed near the irradiated samples. For the first method, the irradiated element was placed in an ionization chamber, and the current was measured with an electroscope. Measurements showed that the ionization produced decreased with the time after the end of x-ray irradiation, but in the case of lead, was measurable 90 minutes after irradiation. A table showing this variation was also given.

An attempt was made by the writer to measure the ionization current produced by metals irradiated with x-rays. The samples were irradiated with 45 kv and with 160 kv x-rays at a distance of about 20 cm from the x-ray tube targets, and for various times. They were then placed in a copper ionization chamber, and the ionization current was measured with a Lindemann electrometer, with a voltage sensitivity of 400 divisions per volt.

[1] G. I. Pokrowski, Phys. Rev. **38,** 925 (1931).

The capacity of the electrometer and chamber was about 12 cm, and the residual ionization in the chamber amounted to about 8.5 ions per cc per second. When a plate of unirradiated lead 12 cm square was placed in the chamber, the electrometer readings were quite irregular, and the total ionization now produced, varied between a maximum of 21.2 ions per cc per second, and a minimum of 17 ions per cc per second. When this same lead was irradiated with 160 kv x-rays for as much as 30 minutes, or with 45 kv x-rays for as much as 45 minutes, the earlier experiments gave a curve of ionization vs. time after irradiation similar to that obtained by Pokrowski, except that the initial ionization was about one half and the slope was steeper. This initial ionization was found to vary considerably, and to be independent of the metal used. However, when improved shielding of the apparatus was provided, no difference whatever could be detected between the ionization produced by unirradiated metals (Pb, Mo, Cu, Al) and that produced by the same metals after irradiation. A particularly troublesome source of heavy ions was traced to the corona produced by other high voltage apparatus in the same room.

If the table that Pokrowski gives records the ionization current in ions per second per square cm of irradiated metal, and not total ionization (which seems probable, but is not definitely stated), then his effect, if it exists, should have been detected easily. In fact, the maximum for lead given by him would have been of the order of 100 times the smallest detectable amount in these experiments. The current measuring device used by Pokrowski is pictured as an electroscope, and, although no mention is made of its current sensitivity, it probably was at least ten times smaller than that of the arrangement used by the writer. Without proper precautions, furthermore, it is possible for very heavy ions to enter the chamber or other sensitive parts of the system, and, according to Hess,[2] may require times of the order of ten minutes to be swept out. It is difficult to explain, however, why Pokrowski should have obtained a relation between the ionization produced, and the atomic number of the metal.

NEWELL S. GINGRICH

Department of Physics, Massachusetts Institute of Technology,

February 1, 1932.

[2] Hess, Electrical Conductivity of the Atmosphere and Its Causes, D. Van Nostrand Company.

BOOK REVIEWS

Numerical Tables of Hyperbolic and Other Functions. J. W. CAMPBELL. Pp. 76, Houghton Mifflin Company, New York, 1929.

This small volume of tables contains a number of novel features which promise to make it very useful. Besides sinh x, cosh x, and tanh x are plotted sinh x/x, (cosh $x-1)/x$, and cosh x/x, of the family of hyperbolic functions. Sinh x and cosh x are tabulated to 4 decimal places for 3 decimal places in the argument from 0 to 3.000 and to 5 significant figures for 2 decimal places in the argument on up to 8.00, a short 7-place table being added for completeness. The next two tables give 4-place tabulations of sin x, cos x, and tan x as functions of the *number* x from $x=0.000$ to 2.000, a table of integral multiples of π being included at the bottom of the page for extension to higher values of x (by means of the formulas for functions of the sum and difference of two numbers). As the author points out these tables are not only useful in practice when x is not given in radian or degree measure, but they should also help the student (who has obviously been kept well in mind in the preparation of these tables) to think of the trigonometric functions just as functions without overstressing their relation to the circle.

Short tables are also included for $\log_{10}x$, $\log_e x$, x^2, x^3, $1/x$ and sin ϕ, cos ϕ, tan ϕ with ϕ in degrees. These last tables are given for decimal intervals of ϕ (intervals of 1/10 degree) instead of in minutes. An appendix on the use of the tables and a glossary of formulas pertaining to the catenary, line transmission problems, and the numerical solution of cubic equations is included.

A set of diagrams of the various functions tabulated might well have been included as well as tables of e^x and e^{-x}, although this last omission is not overly serious because sinh x and cosh x are tabulated symmetrically on opposite (facing) pages (as are sin x and cos x). The make-up of the tables is good, the columns being well set apart, while the eye is not distracted by numerous side columns and cross lines. A single table of proportional parts is glued to the back cover so that it can be opened out and serves at once all of the tables in the book. Since all of the tables are based on the decimal system the student will now find the table of proportional parts of some real value.

E. L. HILL
University of Minnesota

Wave-Length Tables for Spectrum Analysis. F. TWYMAN AND D. M. SMITH. Pp. xi+180. Adam Hilger Ltd., London, Second Edition, 1931. Price, 14s. 9d.

Since the time when the identification of elements by their radiations consisted almost wholly of the practice of examining the colors of heated salts in a bunsen flame, the art of spectrum analysis has made giant strides. It now makes use of the most exact data of spectroscopy. Much of this data has been prepared for the purpose of testing and establishing theories concerning atomic and molecular structure; while a great deal of it consists of wave-length standards for the use of the astrophysicist, the physicist, and the chemist. Hence it is scattered throughout the literature of those sciences. With the recognition of spectrum analysis as a powerful tool for quantitative as well as qualitative determination of the composition of substances, the segregation of this scattered material is imperative. The task has been well performed by the authors, both of whom are eminent in the field of spectrum analysis.

It is a pity that the same volume could not also contain complete information about the practice of spectrum analysis, and we may hope that such a combined volume will come later. The present book contains a few details of technique so that it is something more than a table of wave-lengths. On the whole, however, the title is appropriate. There is a brief though satisfactory introduction to the terminology of spectra, and some reference to simple series relations. The tables include quite complete lists of secondary standards of wave-lengths, and lists of

"distinctive" and "sensitive" lines most useful to the analyst. There is a sufficient bibliography, including full references to the valuable work of de Gramont. The book is a useful addition to the library of the spectroscopist and is invaluable to the technical man in this field.

GEORGE S. MONK
University of Chicago

Fundamentals of Electricity and Magnetism. LEONARD B. LOEB, Pp. xx+432. John Wiley and Sons, 1931. Price $4.00.

This work covers in a manner suitable for students who have studied elementary physics and possess the elements of the calculus the general field of electricity and magnetism together with many important aspects of modern atomic and electrical theory which are germane *to* but not usually included *in* such accounts as are now available. In addition the main subject matter of the book is prefaced by a thoroughly readable and carefully written survey of the historical development of the science from earliest times to the quantum and wave mechanics period.

The scope of this work can best be indicated by a brief mention of the matters treated. It begins with the facts and laws of magnetism, followed by six chapters on electric currents and the laws of electric circuits (steady currents). Then comes electrolysis, a qualitative account of chemical cells and storage batteries and a short chapter on thermoelectricity. It is to be regretted that the treatment of chemical cells could not have been of a more quantitative nature so as to include a brief discussion of Nernst's equation and allied matters.

Electrostatics has quite properly—from the pedagogic point of view at least—been postponed to the middle of the book where it sharpens concepts already introduced and furnishes the background of definitions and ideas necessary for the subsequent treatment of transient phenomena and oscillating circuits.

There follows an account of electromagnetics and induced currents, properties of the magnetic circuit, a brief account of simple dynamos and motors (not intended to be adequate for engineers) and two chapters on alternating currents, including the simple uncoupled oscillating circuit.

Finally comes a condensed sketch of the discharge of electricity in gases, its bearing on the electron and atomic structure and a final (and long) chapter on photoelectricity and thermionic effects. This latter includes a clear discussion of thermionic rectifiers and three electrode tubes used in amplifying and oscillating circuits and closes with a brief mention of piezo and magnetostrictive oscillators. Over thirty pages of problems and exercises collected in an appendix greatly increase the usefulness of the book as a text.

The rather large number of misprints through the book fortunately lead to no ambiguity. The reviewer feels that in the question of dimensional formulae a good deal of trouble on the part of students is eliminated by the conventional method of writing all dimensional formula in large square brackets. Thus the confusion of interpreting symbols such as f both as force and a shorthand notation for "the dimensions of force" is removed. (See for example page 65 et seq.) Again when the advantages of the idea of potential (page 186) are being emphasized—the force at a point is defined as the "rate of change of potential with the distance at the point in question," which is quite correct but conceals the fact that from this standpoint we must know the *direction* of the force before we know in what direction to differentiate (a fact which is known only in the simplest cases). It would surely not be much more difficult to insist on the force in *any* direction as being the derivative of potential in *that* direction—the force at the point being along the direction of most *rapid* change of V with distance.

But these are criticisms of minor import and do not detract from the fact that this work is a very sound and up-to-date addition to the intermediate books on electricity and magnetism —and should fill a want of several years standing.

F. H. CRAWFORD
Harvard University

Radioelements and Isotopes: Chemical Forces and Optical Properties of Substances. KASIMIR FAJANS. The George Fisher Baker Non-Resident Lectureship in Chemistry at Cornell University. Pp. 125+x, figs. 34. The McGraw-Hill Book Company, Inc., New York, 1931. Price $2.50.

This volume presents much of the subject matter of Professor Fajans' lectures at Cornell University during the Spring of 1930 as George Fisher Baker Non-Resident Lecturer in Chemistry. Following an introductory lecture, on the development of views regarding the nature of chemical forces, the book is divided into two parts. Part I is concerned with radioelements and isotopes. Since much of this subject is given in previous works the author limits his discussion to origin of the actinium series and the stability of isotopes. Part II is longer than Part I and is devoted to the subject of chemical forces and optical properties. This section contains seven chapters dealing with the development of this subject. Here the author in discussing the subject presents briefly but concisely the results of his own researches in this important field of chemistry. To anyone wishing to become acquainted with this realm of investigation this book should serve as an excellent introduction.

L. H. Reyerson
University of Minnesota

CURRENT LITERATURE OF PHYSICS

Extracts from the Tables of Contents of Some of our Contemporaries

PROCEEDINGS OF THE ROYAL SOCIETY

SERIES A. VOL. 134, NO. A825 — JANUARY 2, 1932

Address of the President, **Sir F. Gowland Hopkins,** at the Anniversary Meeting, November 30, 1931 533

Sorption of Carbon Bisulphide at Low Pressures by Activated Charcoals. By **A. J. Allmand, and J. L. Lizius** 554

Isotopic Constitution and Atomic Weights of Caesium, Strontium, Lithium, Rubidium, Barium, Scandium and Thallium. By **F. W. Aston** 571

Slow Combustion of Methane. By **W. A. Bone and R. E. Allum** 578

Formation of Methyl Alcohol and Formaldehyde in the Slow Combustion of Methane at High Pressures. By **D. M. Newitt and A. E. Haffner** 591

Third Spark Spectrum of Arsenic (As IV). By **K. R. Rao** 604

Field of Force and the Form of the Carbon Dioxide Molecule. Part I Thermal Diffusion in Mixtures Containing Carbon Dioxide. By **T. L. Ibbs and A. C. R. Wakeman** 613

Field of Force and the Form of the Carbon Dioxide Molecule. Part II. Variation of the Viscosity of Carbon Dioxide with Temperature. By **T. L. Ibbs and A. C. R. Wakeman** 628

Field of Force and the Form of the Carbon Dioxide Molecule. Part III. Form of the Carbon Dioxide Molecule. By **T. L. Ibbs and A. C. R. Wakeman** 636

Intensity Distribution in the Band Spectrum of Helium (He_2) for the Region λ6100-λ6500. By **C. W. W. Read** 643

Further Investigations with a Wilson Chamber. II. Range and Velocity of Recoil Atoms. By **P. M. S. Blackett and D. S. Lees** 658

Distribution of Energy in the β-Ray Spectrum of Radium E. By **F. C. Champion** 672

Crystalline Vitamin D. By **F. A. Askew, R. B. Bourdillon, H. M. Bruce, R. K. Callow, J. St. L. Philpot and T. A. Webster** (Abstract) 682

NATURE

VOL. 129, NO. 3244 — JANUARY 2, 1932

Letters to the Editor:

Absolute Values of X-ray Wave-lengths and the Fundamental Atomic Constants. By **Manne Siegbahn and Martin Söderman** 21

Experimental Proof of the Spin of the Photon. By **Sir C. V. Raman and S. Bhagavantam** 22

Wave-length of the Green Auroral Line Determined by the Interferometer. By **L. Vegard** 23

Some effects of Turbulent Pressure. By **O. F. T. Roberts** 23

Ether-Drift and Gravity. By **R. Tomaschek and W. Schaffernicht** 24

Diamagnetism of Liquid Mixtures. By **John Farquharson** 25

General Expression for Intensity of Hydrogen Lines. By **L. McLean** 25

NATURE

VOL. 129, NO. 3245 — JANUARY 9, 1932

Letters to the Editor:

Anomalous Scattering of α Particles by Hydrogen and Helium. By **H. M. Taylor** 56

Nuclear Spin and Hyperfine Structure in Band Spectra. By **E. Hulthén** 56

Significance of Velocity Measurements in Relation to the Benzene Substitution Problem. By **D. H. Peacock** 57

Hall Effect and Superconductivity. By **I. Kikoin and Boris Lasarew** 57

Liquid Drops on the Same Liquid Surface. By **William Hughes and Alun R. Hughes** 59

New Method of Investigating the Modes of Vibration of Quartz Crystals. By **J. A. Strong** 59
Electrostatic Explanation of the Phenomenon of Flotation. By **B. Kamienski** 59
Nature of Liquids. By **L. G. Carpenter** 60
Raman Band of Water. By **G. Bolla** 60

NATURE

VOL. 129, NO. 3246 JANUARY 16, 1932

Electron Optics. By **G. P. Thomson** 81

Letters to the Editor:

Velocities of the Spiral Nebulae. By **W. D. MacMillan** 93
L-Discontinuities in X-ray Absorption. By **E. Olsson** 94
Atomic Moments of Some Metals. By **Pierre Weiss** 95
Diamagnetism of Liquid Mixtures. By **James F. Spencer and C. G. Trew** 95
Structure and Development of Temperature Inversions in the Atmosphere. By **Sobhag Mal, S. Basu and B. N. Desai** 97

ZEITSCHRIFT FÜR PHYSIK

VOL. 73, NOS. 9 AND 10 JANUARY 8, 1932

On the Problem of Chemical Combination as applied to Organic Chemistry. By **F. Hund** 565
Methods of Measuring Atomic Disintegration. By **G. Hoffmann** 578
New Investigations on the Actions of Electrolytic Valves. III, Dielectric Constant of Al_2O_3 Resisting Layer. IV, Conduction Mechanism in the Resisting Layer of the Valve Metal. By **A. Güntherschulze and Hans Betz** 580
On an Effect which Appears in Ferromagnetic Materials in Alternating Electromagnetic Fields. By **A. Esau and H. Kortum** 602
On Terms of the Krypton Arc Spectrum. By **W. Gremmer** 620
Influence of Foreign Substances on the Cohesion Limits and the Ultra-Microscopic Structure of Synthetic Rock Salts. By **Alfred Edner** 623
Polarization of Mercury Fluorescence in Relation to the Line Hyperfine Structure. By **V. v. Keussler** 649
Investigations of the Detection of Raman Effects Near Metal Electrodes. By **F. W. Loomis and J. G. Winans** 658
On the Production of the Continuous Spectra by Cathode-Ray Bombardment. V. Remarks on the Investigations in a Lilienfeld Tube and Tests of other Tubes. By **Willi Cohn** 662
On the Barometric Coefficient of Hard Ultra-Radiation. By **W. Messerschmidt and W. S. Pforte** 677
Appearance of Metal-Noble Gas Bands in the Positive Columns of Electrical Discharges. By **H. Krefft and R. Rompe** 681
Influence of Supporting Layer on the Temperature Coefficient of the Electric Resistance of Thin Platinum Films. By **G. Baintner** 691
Piezoelectric Properties of Rochelle Salts Crystals. By **Frau R. D. Schulwas-Sorokin** 700
Concerning the Association in Strong Electrolytes. By **Johannes Zirkler** 707

DIE NATURWISSENSCHAFTEN

VOL. 20, NO. 3 JANUARY 15, 1932

Visibility of Ultraviolet Light. By **E. Merker** 41

Original Contributions:

Electron Microscope. By **E. Brüche** 49
On Anomalies in the Fine Structure of the Self Absorption of Spectral Lines. By **E. Lau and O. Reichenheim** 49

Discussions:

On the Bending of Light in the Gravitational Field of the Sun 50
Relations between the Eruptions on the Sun and Earth Magnetic Storms with Reference to the Aurora. By **W. Grotrian** 55

PHYSIKALISCHE ZEITSCHRIFT

VOL. 33, NO. 1 — JANUARY 1, 1932

Band Spectroscopy and Photo-Chemistry. By **R. Mecke** 1

Original Contributions:

Theory of the Magnetization Curve of Isotropic Ferromagnetic Bodies in Moderate and Strong Fields. By **R. Gans** 15

Action of Polarized Light on the Resisting Layer of the Photo Cell. By **L. Bergmann** 17

Absorption of Dipole Liquids and Electrolytic Solutions in the Region of Short Electric Waves. By **J. Malsch** 19

Application of Debye Electrolytic Theory to Concentrated Solutions. By **G. Scatchard** 22

Elasticity and Rigidity of Thin Quartz Fibres. By **O. Reinkober** 32

Influence of Quantization on the Calculation of the Virial Coefficient. By **J. G. Kirkwood** 39

Use of Analogies in the Flow of Electricity in Gases. By **H. Mache** 43

Discussions:

Absorption Discontinuity in the *L*-Bands of Heavy Elements. By **H. Küstner** 46

On the Question of Normal Atomic Weights. By **R. Mecke** 49

ANNALEN DER PHYSIK

VOL. 12, NO. 4 — JANUARY 8, 1932

On the Complete Differential Secondary Radiation in Air for Electrons of Moderate Velocity. By **E. Kipphan** 401

Investigation of the Motion of Electrons in Flames. By **Karl Vogt** 433

Behaviour of Electric Waves in the Passage through Ionized Gases. By **Werner Hasselback** 477

Determination of the Specific Charge of the Electron from Velocity Measurements. By **Fritz Kirchner** 503

Extension of the Spectra of Al II, Mg I, Be I, and Al I. By **E. Paschen** 509

On the Sound Radiation from a Circular Vibrating Membrane with Nodal Lines. By **H. Stenzel** 528

ANNALEN DER PHYSIK

VOL. 12, NO. 5 — JANUARY 18, 1932

Angular Distribution in the Scattering of Slow Electrons by Gas Molecules. By **C. Ramsauer and R. Kollath** 529

Isothermal Measurements of Hydrogen between Zero and 100°C up to 1000 Atmospheres. By **A. Michels, G. P. Nijhoff and A. J. J. Gerver** 562

Gas Discharge at very High Frequencies. By **Lothar Rohde** 569

Remarks on the Range and Ionization Powers of H' and α-Rays. By **E. Rüchardt** 600

Contributions to Geometric Electron Optics. By **M. Knoll and E. Ruska** 607

BUREAU OF STANDARDS JOURNAL OF RESEARCH

VOL. 8, NO. 3 — MARCH, 1932

Heat of Vaporization of Water at 50°, 70°, and 90°C. By **E. F. Fiock and D. C. Ginnings** .. 416

Note on International Comparison of X-ray Standards. By **Lauriston S. Taylor** 417

Course of the Oxidation of the Aldose Sugars by Bromine Water. By **H. S. Isbell and C. S. Hudson** 418

Fluctuations of the Rate of Emissions of α-particles for Weak Sources and Large Solid Angles. By **L. F. Curtiss** 419

Gases Obtained from Commercial Feldspars Heated in Vacuo. By **G. R. Shelton and H. H. Holscher** 420

Experiments on the Emission and Absorption of Radiation by Metallic Silver. By **F. L. Mohler** 421

Accelerated Weathering Tests of Soldered and Tinned Sheet Copper. By **Peter R. Kosting** 422
Chromium Oxide and the Vanadium Oxide Band Spectra. By **W. F. C. Ferguson** 423
Air Density Corrections for X-ray Ionization Chambers. By **Lauriston S. Taylor and George Singer** 424
Arc Spectrum of Phosphorous. By **C. C. Kiess** 425
Some Physical Properties of Levulose and its Estimation by Copper Reduction Methods. By **Richard F. Jackson and Joseph A. Mathews** 426

JOURNAL OF THE FRANKLIN INSTITUTE

VOL. 213, No. 2 FEBRUARY, 1932

Solution of Steady-State Problems in Dielectric, Magnetically Permeable, and Conducting Media, with Special Reference to Mathematical Analogies between Magnetic Problems and Current-Flow Problems. By **W. F. G. Swann** 155
Effect of Electron Attachment on the Ion Mobility Curves in the Zeleny Air Blast Method of Ion Mobility Measurement. By **Leonard B. Loeb and Norris E. Bradbury** 181

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA

VOL. 22, No. 2 FEBRUARY, 1932

Problem of Motion Picture Projection from Continuously Moving Film. By **Fordyce Tuttle and Charles. D. Reid** 39
"Depressed" Beam of the Motor Car Headlight. By **L. W. Taylor** 65
Chromaticity Sensibility to Stimulus Differences. By **Deane B. Judd** 72

THE REVIEW OF SCIENTIFIC INSTRUMENTS

VOL. 3, No. 2 FEBRUARY, 1932

Notes on Radiation Thermopiles. By **John Strong** 65
Oscillator for the Crystal of an X-ray Spectrograph. By **Gerald W. Fox** 71
Tellurium-Bismuth Vacuum Radiation Thermocouple. By **C. Hawley Cartwright** 73
New Type of Radiometer. By **Alexander L. M. Dingee** 80
Stroboglow. By **L. R. Quarles** 85
Improved High Resistance Unit. By **H. C. Rentschler and D. E. Henry** 91
Sensitive Flexible Thermostat. By **J. R. Roebuck** 93